KNOWLEDGE ENGINEERING

Building Cognitive Assistants for Evidence-Based Reasoning

This book presents a significant advancement in the theory and practice of knowledge engineering, the discipline concerned with the development of intelligent agents that use knowledge and reasoning to perform problem-solving and decision-making tasks. It covers the main stages in the development of a knowledge-based agent: understanding the application domain, modeling problem solving in that domain, developing the ontology, learning the reasoning rules, and testing the agent. The book focuses on a special class of agents: cognitive assistants for evidence-based reasoning that learn complex problem-solving expertise directly from human experts, support experts, and nonexperts in problem solving and decision making, and teach their problem-solving expertise to students.

A powerful learning agent shell, Disciple-EBR, is included with the book, enabling students, practitioners, and researchers to develop cognitive assistants rapidly in a wide variety of domains that require evidence-based reasoning, including intelligence analysis, cybersecurity, law, forensics, medicine, and education.

Gheorghe Tecuci (PhD, University of Paris-South, July 1988, and Polytechnic Institute of Bucharest, December 1988) is Professor of Computer Science and Director of the Learning Agents Center in the Volgenau School of Engineering of George Mason University, Member of the Romanian Academy, and former Chair of Artificial Intelligence in the Center for Strategic Leadership of the U.S. Army War College.

Dorin Marcu (PhD, George Mason University, 2009) is Research Assistant Professor, as well as Senior Software and Knowledge Engineer, in the Learning Agents Center, Volgenau School of Engineering, George Mason University.

Mihai Boicu (PhD, George Mason University, 2003) is Associate Professor of Information Sciences and Technology, and Associate Director of the Learning Agents Center, Volgenau School of Engineering, George Mason University.

David A. Schum (PhD, Ohio State University, 1964) is Emeritus Professor of Systems Engineering, Operations Research, and Law, as well as Chief Scientist of the Learning Agents Center at George Mason University. He is also Honorary Professor of Evidence Science at University College London.

Knowledge Engineering

Building Cognitive Assistants for Evidence-Based Reasoning

GHEORGHE TECUCI
George Mason University

DORIN MARCU
George Mason University

MIHAI BOICU
George Mason University

DAVID A. SCHUM
George Mason University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

One Liberty Plaza, New York, NY 10006

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/9781107122567

First published 2016

Printed in the United States of America by Sheridan Books, Inc.

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Tecuci, Gheorghe, author. | Marcu, Dorin, author. | Boicu, Mihai, author. |
Schum, David A., author.
Title: Knowledge engineering: building cognitive assistants for evidence-based reasoning /
Gheorghe Tecuci, George Mason University, Dorin Marcu, George Mason University,
Mihai Boicu, George Mason University, David A. Schum, George Mason University.
Description: New York NY : Cambridge University Press, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2015042941 | ISBN 9781107122567 (Hardback : alk. paper)
Subjects: LCSH: Expert systems (Computer science) | Intelligent agents (Computer software) | Machine learning |
Artificial intelligence | Knowledge, Theory of–Data processing.
Classification: LCC QA76.76.E95 T435 2016 | DDC 006.3/3–dc23 LC record available at
http://lccn.loc.gov/2015042941

ISBN 978-1-107-12256-7 Hardback

Contents

Preface

BOOK PURPOSE

This is a book on knowledge engineering, the discipline concerned with the development of intelligent agents that use knowledge and reasoning to perform problem-solving and decision-making tasks. The book covers the theory and practice of the main stages in the development of a knowledge-based agent: understanding the application domain, modeling problem solving in that domain, developing the ontology, learning the reasoning rules, and testing the agent. However, it does this by focusing on a special class of agents: cognitive assistants that learn complex problem-solving expertise directly from human experts, support experts, and nonexperts in problem solving and decision making and teach their problem-solving expertise to students. These are learning agents that are taught by their users in ways that are similar to how a student, an apprentice, or a new collaborator is taught, through problem-solving examples and explanations and by supervising and correcting their behavior. Because such agents learn to replicate the problem-solving behavior of their users, we have called them Disciple agents.

This book presents a significant advancement in the theory and practice of knowledge engineering, where many tasks are performed by a typical computer user and a learning agent, with only limited support from a knowledge engineer. To simplify further the development of the cognitive assistants by typical users, we have focused on the development of cognitive assistants for evidence-based reasoning. Evidence-based reasoning is at the core of many problem-solving and decision-making tasks in a wide variety of domains, including intelligence analysis, cybersecurity, law, forensics, medicine, physics, chemistry, history, archaeology, education, and many others. Nevertheless, the last part of the book presents Disciple agents for applications that did not involve evidence-based reasoning.

Because knowledge engineering is a practical activity, it is best learned by doing. Therefore, this book presents the theory and methodology of developing cognitive assistants in conjunction with a practical tool, Disciple-EBR, a learning agent shell for evidence-based reasoning (EBR). Consequently, each chapter typically contains a theoretical part presenting general concepts and methods, a methodological part with guidelines on the application of the methods, and a practical part on the actual employment of these methods with Disciple-EBR. It also includes project assignments and review questions.

This book addresses issues of interest to a large spectrum of readers from academia, research, and industry. We have used drafts of this book in our computer science courses on knowledge engineering, expert systems, and knowledge-based agents, at both undergraduate and graduate levels, because it covers the theory and practice of the main stages in the development of knowledge-based agents. We have also used some parts of the book in introductory courses on artificial intelligence, and other parts in the courses on knowledge acquisition and machine learning. These are all examples of courses where this book could be used.

Researchers in knowledge engineering will find in this book an integrated approach that advances the theory and practice in the field. We believe that further research and development of this approach will enable typical computer users to develop their own cognitive assistants without any knowledge engineering assistance. Thus, non–computer scientists will no longer be only users of generic programs developed by others (such as word processors or Internet browsers), as they are today, but also agent developers themselves. They will be able to train their personal assistants to help them with their increasingly complex tasks in the knowledge society, which should have a significant beneficial impact on their work and life.

Practitioners that develop various types of knowledge-based systems will find in this book a detailed, yet intuitive, presentation of an agent development methodology and tool, as well as several case studies of developing intelligent agents that illustrate different types of agents that are relevant to a wide variety of application domains. In fact, a complementary book, *Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots* (Tecuci et al., 2016), presents the practical application of Disciple-CD, an agent developed with Disciple-EBR for intelligence analysis problems.

BOOK CONTENTS

Here is a route or map we will follow in the learning venture you will have with the assistance of Disciple-EBR. Chapter 1 is a general introduction to the topics that form the basis of this book. It starts with the problem of understanding the world through evidence-based reasoning. It then presents abductive reasoning, five different conceptions of probability (enumerative, subjective Bayesian, Belief Functions, Baconian, and Fuzzy), and how deductive, abductive, and inductive (probabilistic) reasoning are used in evidence-based reasoning. After that, it introduces artificial intelligence and intelligent agents, and the challenges of developing such agents through conventional knowledge engineering. Afterward, it introduces the development of agents through teaching and learning, which is the approach presented in this book.

Chapter 2 is an overview of evidence-based reasoning, which is a focus of this book. It starts with a discussion of the elements that make evidence-based reasoning an astonishingly complex task. It then introduces a systematic approach that integrates abduction, deduction, and induction to solve a typical evidence-based reasoning task, using intelligence analysis as an example. Finally, it shows the application of the same approach to other

evidence-based reasoning tasks in cybersecurity, geospatial intelligence, and critical thinking education.

Chapter 3 is an overview of the main methodologies and tools for the design and development of knowledge-based agents. It first presents the conventional approach of developing such agents by a knowledge engineer working with a subject matter expert. It then introduces different types of agent shells, which are the typical tools for building agents, and discusses the use of foundational and utility ontologies. The more advanced of these tools, the learning agent shells, are at the basis of a new and more powerful approach to agent design and development, an overview of which is presented in the second part of Chapter 3. This learning-based approach is illustrated with the development of a cognitive assistant for assessing a potential PhD advisor for a student, an example that is used in the following chapters to present in detail each of the main agent development stages.

Chapter 4 presents the modeling of the problem-solving process through analysis and synthesis, which is the most challenging task in the development of a knowledge-based agent. This chapter starts with a more general version of this process for any type of problems, which is used in the Disciple agents presented in Chapter 12. After that, the chapter presents an easier, customized version for evidence-based hypothesis analysis, which is used in the chapters that follow. Chapter 4 also introduces an ontology of evidence and presents the modeling of the believability assessment for different types of evidence.

Chapters 5 and 6 present the representation of knowledge through ontologies, as well as their design and development. They discuss the representation of the objects from the agent's application domain and their relationships. These chapters also discuss the basic reasoning operations of transitivity, inheritance, and matching. They cover several approaches to concept elicitation, as well as a systematic approach to modeling-based ontology specification. The chapters also address the complexity of ontology maintenance.

Chapter 7 presents the agent's reasoning based on ontologies and rules, in the context of a production system architecture. It starts with defining the representation of complex ontology-based concepts and the use of these concepts in the reasoning rules. It defines the reduction and syntheses rules in general, and the special case of these rules for evidence-based reasoning. It then presents the process of problem solving through analysis and synthesis, which is accomplished through rule and ontology matching. This chapter also introduces the representation of and reasoning with partially learned concepts, features, hypotheses, and rules.

Chapter 8 starts with a general introduction to machine learning and to several learning strategies that are most relevant for knowledge engineering. It continues with the definition of the generalization and specializations of concepts through the use of inductive rules. After that, the chapter defines the basic operations of minimal and maximal generalizations and specialization of concepts, which are at the basis of rule learning and refinement discussed in Chapters 9 and 10. The chapter ends with the presentation of a formal definition of generalization.

Chapter 9 presents the mixed-initiative rule learning method that enables an agent to learn a general rule from a specific example of a

reasoning step. It starts with an overview of the integration of modeling, learning, and problem solving and an illustration of this process. The chapter then introduces the rule learning problem and method. After that, it presents in detail the phases of rule learning, such as the mixed-initiative understanding of the example, and its analogy-based generalizations, which result in a minimal and a maximal generalization forming the plausible version space of the rule. This chapter also discusses the learning of rules involving functions and relational operators.

The refinement of a partially learned rule is presented in Chapter 10. After introducing the incremental rule refinement problem, this chapter presents the refinement of the rule based on positive examples, and then the refinement based on negative examples. This may result in a very complex rule characterized by a main applicability condition and several "except-when" conditions, which capture the reasoning of a subject matter expert. There are various refinement strategies, depending on the position of the examples with respect to the bounds of the rule's conditions and also on the possible explanations of these examples, but in all cases, they involve simple interactions with the subject matter expert. Changes in the ontology require the learned rules to be updated, the corresponding method being presented and illustrated in the second part of this chapter. The chapter ends with a characterization of rule learning and refinement, which enable a non–computer scientist to teach an agent.

The chapters dedicated to the individual phases of agent development end with Chapter 11, which discusses the abstractions of individual hypotheses and of the reasoning tree to facilitate its browsing, understanding, and further development by the end-user.

As previously indicated, this book focuses on the development of agents for evidence-based reasoning tasks. However, the presented theory and methodology are also applicable to other types of agents. These agents have been developed with learning agent shells representing previous implementations of the Disciple theory and methodology. Four of these agents are presented in Chapter 12. Disciple-WA is an agent that uses expert knowledge from military engineering manuals to develop alternative plans of actions that a military unit can use to work around (WA) damages to a transportation infrastructure, such as a damaged, destroyed, or mined bridge, road, or tunnel. The goal is to find the quickest way for the military unit to bypass the encountered obstacle. There were several cases where the Disciple-WA agent generated better solutions than those of the human expert who evaluated the developed systems, as well as cases where the agent generated new solutions that this expert did not consider.

Disciple-COA is an agent trained by a military expert to critique courses of action (COA) with respect to the principles of war and the tenets of Army operations. It receives as input the description of a course of action and returns a list of strengths and weaknesses of various levels, together with their justifications. A remarkable feature of this agent is that it was judged to exceed the performance of the subject matter experts that defined the problems for its evaluation.

Disciple-COG is an agent that was trained to identify and assess the center of gravity (COG) candidates of the opposing forces in a military scenario. Correctly identifying the centers of gravity of the opposing forces

is of highest importance in any conflict, and Disciple-COG has been used for many years in courses at the U.S. Army War College, as well as at other military institutions, to teach students how to perform a strategic center-of-gravity analysis of a scenario of interest.

The last agent described in Chapter 12 is Disciple-VPT (Virtual Planning Team). Disciple-VPT consists of virtual planning experts that collaborate to develop plans of actions requiring expertise from multiple domains. They are assembled from an extensible library of such agents. The basic component of Disciple-VPT is the Disciple-VE learning agent shell that can be taught how to plan directly by a subject matter expert. Copies of the Disciple-VE shells can be used by experts in different domains to rapidly populate the library of virtual experts (VEs) of Disciple-VPT.

The learning-based approach to knowledge engineering presented in this book illustrates the application of several design principles that are useful in the development of cognitive assistants in general. Therefore, as a conclusion, Chapter 13 summarizes these principles, which are illustrated throughout this book.

The book also includes several appendices that summarize important aspects from the chapters: the list of the knowledge-engineering guidelines for each of the main stages of agent development, the list of operations of Disciple-EBR, and the list of the hands-on exercises. Answers to selected questions from each chapter are made available to the instructors.

BACKGROUND

The learning-based knowledge-engineering theory, methodology, and tools presented in this book are the result of many years of research and experimentation that produced increasingly more general and powerful versions. During this evolution, one may identify four main stages. The first stage corresponds to the PhD work of Gheorghe Tecuci presented in his thesis "Disciple: A Theory, Methodology and System for Learning Expert Knowledge" (Thèse de Docteur en Science, Université de Paris-Sud, July 1988). The main emphasis of that work was on rule learning. The developed methods are among the first multistrategy approaches to learning that later grew into the subfield of multistrategy machine learning (Tecuci, 1993; Michalski and Tecuci, 1994). They also are among the first attempts to integrate machine learning and knowledge acquisition (Tecuci et al., 1994; Tecuci and Kodratoff, 1995). This work benefited from the collaboration of Yves Kodratoff and the support of Mihai Drăgănescu. It was done with the support of the Romanian Research Institute for Informatics, the Romanian Academy, and the French National Center for Scientific Research.

The second stage in the evolution of the Disciple approach is presented in the book *Building Intelligent Agents: An Apprenticeship Multistrategy Learning Theory, Methodology, Tool and Case Studies*, published by Academic Press in 1998. It includes an improvement of Disciple's multistrategy learning methods and their extension with guided knowledge-elicitation methods. It also includes ontology development tools and illustrations of the application of the Disciple rule-learning approach to a variety of domains, such as teaching of higher-order thinking skills in history and

statistics, engineering design, and military simulation. This work benefited from the collaboration of several of Tecuci's PhD students, particularly Thomas Dybala, Michael Hieb, Harry Keeling, and Kathryn Wright, and it was partially supported by George Mason University, the National Science Foundation, and the Defense Advanced Research Projects Agency.

The third stage in the evolution of the Disciple approach is represented by the Disciple agents described in Chapter 12 of this book. This represents a significant extension of the Disciple approach, with methods for problem solving through analysis and synthesis, modeling of the problem solving process, ontology development, and scenario elicitation. At this stage, Disciple has become a complete, end-to-end agent development methodology that has been applied to develop powerful agents, such as Disciple-COG, used for a period of ten years in courses at the U.S. Army War College and at other institutions. This work also benefited from the collaboration of Tecuci's students, first of all Mihai Boicu and Dorin Marcu, and also Michael Bowman, Vu Le, Cristina Boicu, Bogdan Stanescu, and Marcel Barbulescu. This work was partially supported by George Mason University, the Air Force Office of Scientific Research, the Air Force Research Laboratory, the Defense Advanced Research Projects Agency, the National Science Foundation, and others.

Finally, the latest stage in the evolution of the Disciple approach is represented by the rest of this book. A main conceptual advancement over the previous stage consists in its extension with a theory of evidence-based reasoning, greatly facilitated by the collaboration of David A. Schum. This resulted in a very powerful theory, methodology, and tool for the development of agents for complex evidence-based reasoning applications, such as intelligence analysis. Two of such agents are TIACRITIS (Teaching Intelligence Analysts Critical Thinking Skills) and Disciple-CD (Disciple cognitive assistant for Connecting the Dots). This work was partially supported by George Mason University and by several agencies of the U.S. government, including the Department of Defense.

Acknowledgments

We are very grateful to the many individuals who, in various ways, supported our research, including Kelcy Allwein, Cindy Ayers, Murray Burke, Douglas Campbell, William Cleckner, Jerry Comello, John Donelan, Jim Donlon, Susan Durham, Keri Eustis, Michael Fletcher, Erin Gibbens, Lloyd Griffiths, David Gunning, Ben Hamilton, Sharon Hamilton, Robert Herklotz, Phillip Hwang, Eric Jones, Alex Kilpatrick, David Luginbuhl, Joan McIntyre, Jean-Michel Pomrade, Michelle Quirk, Peter Rocci, William Rzepka, Kimberly Urban, Joan Vallancewhitacre, and Ben Wible.

We also want to thank Heather Bergman, who invited us to write this book for the Cambridge University Press, as well as to senior editor Lauren Cowles, who is a great professional to work with.

About the Authors

Gheorghe Tecuci (PhD, Université de Paris-Sud, July 1988, and Polytechnic Institute of Bucharest, December 1988) is Professor of Computer Science and Director of the Learning Agents Center in the Volgenau School of Engineering of George Mason University, Member of the Romanian Academy, and former Chair of Artificial Intelligence in the Center for Strategic Leadership of the U.S. Army War College. He has published around two hundred papers, including eleven books, with contributions to artificial intelligence, knowledge engineering, cognitive assistants, machine learning, evidence-based reasoning, and intelligence analysis. He has received the U.S. Army Outstanding Civilian Service Medal (for "groundbreaking contributions to the application of artificial intelligence to center of gravity determination") and the Innovative Application Award from the American Association for Artificial Intelligence.

Dorin Marcu (PhD, George Mason University, 2009) is Research Assistant Professor, as well as Senior Software and Knowledge Engineer, in the Learning Agents Center, Volgenau School of Engineering, George Mason University. He has published more than forty papers, including five books, with contributions to adaptive user interfaces, mixed-initiative human–computer interaction, and cognitive assistants. He has received the Innovative Application Award from the American Association for Artificial Intelligence.

Mihai Boicu (PhD, George Mason University, 2002) is Associate Professor of Information Sciences and Technology and Associate Director of the Learning Agents Center, in the Volgenau School of Engineering of George Mason University. He has published over ninety papers, including five books, with contributions to problem solving and multistrategy learning in dynamic and evolving representation spaces, mixed-initiative interaction, multi-agent systems architecture, collaboration and coordination, abstraction-based reasoning, knowledge representation, and knowledge acquisition. He has received the Innovative Application Award from the American Association for Artificial Intelligence.

David A. Schum (PhD, Ohio State University, 1964) is Emeritus Professor of Systems Engineering, Operations Research, and Law, as well as Chief Scientist of the Learning Agents Center at George Mason University, and Honorary Professor of Evidence Science at University College London.

He has followed a career-long interest in the study of the properties, uses, discovery, and marshaling of evidence in probabilistic reasoning. Dr. Schum has published more than one hundred papers in a variety of journals and eight books, including *The Evidential Foundations of Probabilistic Reasoning, Analysis of Evidence, Evidence and Inference for the Intelligence Analyst,* and *Probabilistic Analysis of the Sacco and Vanzetti Evidence,* being recognized as one of the founding fathers of the Science of Evidence.