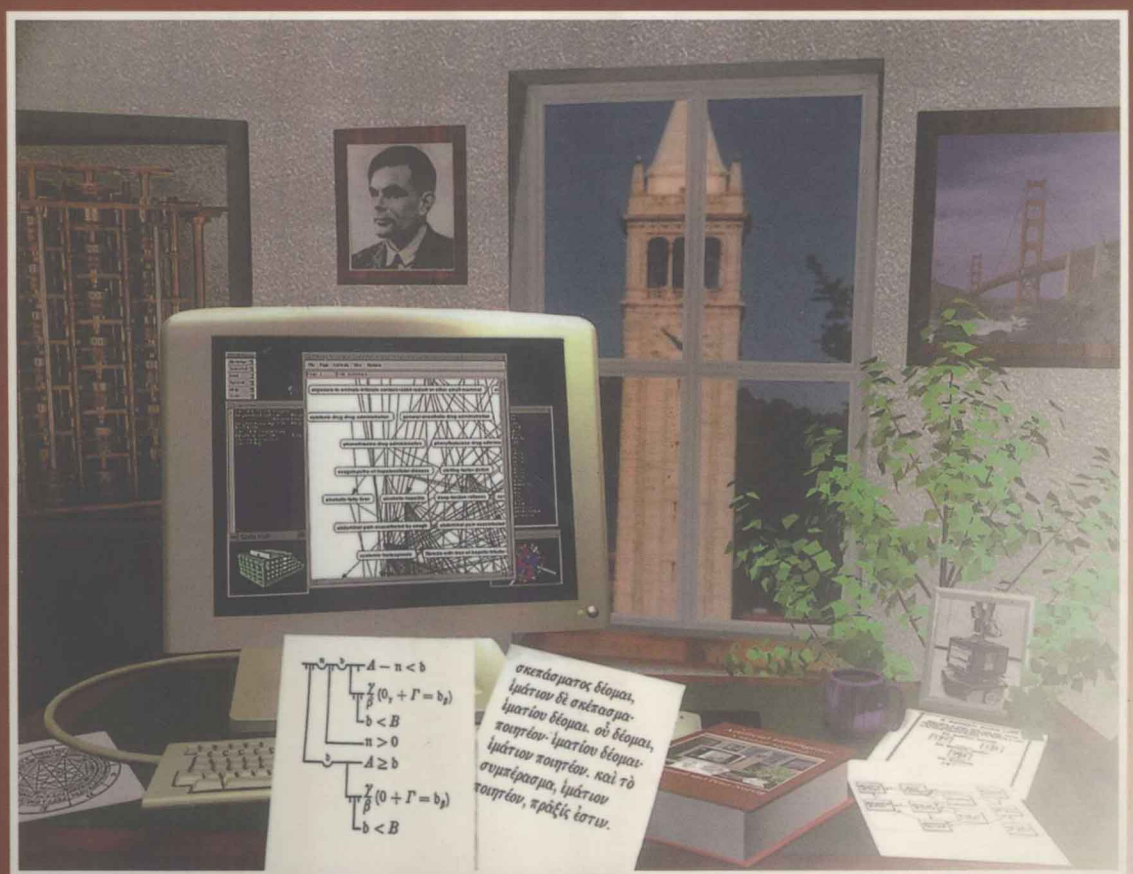
国外著名高等院校
信息科学与技术优秀教材

人工智能

——一种现代方法

Artificial Intelligence: A Modern Approach



Stuart Russell · Peter Norvig

英文版

Prentice Hall
Pearson Education
出 版 集 团

人民邮电出版社
POSTS & TELECOMMUNICATIONS PRESS

国外著名高等院校信息科学与技术优秀教材

人工智能——一种现代方法

(英文版)

Artificial Intelligence: A Modern Approach

Stuart Russell

Peter Norvig

人民邮电出版社 Prentice Hall Pearson Education 出版集团

图书在版编目(CIP)数据

人工智能：一种现代方法/（英）拉塞尔（Russell，S.）（美）诺文（Norvin，P.）著.
—北京：人民邮电出版社，2002.4
国外著名高等院校信息科学与技术优秀教材
ISBN 7-115-10202-3
I. 人… II. ①拉…②诺… III. 人工智能—高等学校—教材—英文 IV. TP18

中国版本图书馆 CIP 数据核字(2002)第 015670 号

版 权 声 明

国外著名高等院校信息科学与技术优秀教材

人工智能 —— 一种现代方法（英文版）

◆ 著 Stuart Russell Peter Norvig
责任编辑 李 际

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
读者热线 010-67180876
北京汉魂图文设计有限公司制作
北京朝阳展望印刷厂印刷
新华书店总店北京发行所经销

◆ 开本：800×1000 1/16
印张：60.25
字数：1 342 千字 2002 年 4 月第 1 版
印数：1-3 000 册 2002 年 4 月北京第 1 次印刷

著作权合同登记 图字：01－2002－1554 号

ISBN 7-115-10202-3/TP・2834

定价：75.00 元

本书如有印装质量问题，请与本社联系 电话：(010)67129223

序　言

我国人工智能（AI）的教学和科研起步于上个世纪 70 年代末，正是我国实施改革开放政策，打开国门的时候。因此，随后不久国外有关的教材就先后引进到国内来，并很快出版了中译本，以满足教学的需要。其中影响较大的有，美国尼尔森的《人工智能原理》（N.J.Nilsson, 1980；中译本，1983），温斯顿的《人工智能》（P.H.Winston, 1977；中译本，1983），以及里奇的《人工智能》（E.Rich, 1983，中译本，1992）等。以后，国内学者也相继编写和出版了自己的书籍，如清华大学出版社出版的《人工智能导论》（林尧瑞等，1989）和《人工智能原理》（石纯一等，1993）等。这些书籍曾先后作为各个高等院校本科生或研究生的教材，为培养和造就我国 AI 科研和教学人才起到了很好的作用。

由于人工智能是一门发展中的学科，尚未形成系统的理论体系，理论和技术均在不断的变化中。同时它又同许多领域相互交叉，形成各自独立的研究分支，内容繁杂，其中包括计算机科学、数学、心理学与认知科学等，这些都给教材的编写带来一定的困难。比如，（1）由于 AI 历史上各个分支领域独立发展，缺乏联系，将各种不同的主题放在一部书中，很容易造成内容零散且互不相关的现象；（2）有的研究分支还不成熟，难以进行系统和深入的介绍；（3）内容更新很快、很容易过时等。

这里向读者推荐的《人工智能——一种现代方法》（S. Russell, P. Norvig）是一部 AI 的教科书。显然，作者认识到编写人工智能教材存在的上述困难。在该书的序言中特别介绍了他们的意图以及所做的努力。正是在他们的努力下，使该书具有以下明显的特色。

（1）作者试图在智能 Agent（译为自主体、智能体等，由于目前还没有合适的译名，就使用原文）的概念框架下，把 AI 中相互分离的领域统一起来。书的主体内容共分为六大部分，即问题求解、知识与推理、合乎逻辑的行动、不确定知识与推理、学习，以及通信、感知与行动。作者通过 Agent 从感知外部环境、到实施行动、并最后对外部环境施加影响的全过程，将这六部分组织起来，形成一个相互联系的整体，使读者对 AI 有一个完整的概念，达到较好的效果。这种组织方式是新颖的，并得到同行的认可，如最近尼尔森编著的《人工

智能——一种新的综合法》（N.J.Nilsson, Artificial Intelligence: A New Synthesis, 1998; 中译本，2000）也采用了类似的组织方法和思路。

（2）人工智能是一门新兴的交叉学科，因此对于什么是 AI，不同的研究者有不同的认识。由于认识上的差异，因此作为介绍 AI 的书籍，也就有了不同的定位和重点。大体说来，AI 具有科学和工程两个方面。人们的研究可以侧重于 Agent 的认知过程，也可以侧重于行为的效果。显然，前者对应于 AI 的科学问题；后者对应于工程实现。同样是研究认知过程，人人可以把机器与人类思维过程是否相似作为评判“智能”的标准，也可以把智能机器是否达到某一个给定的客观性能（如合理性，rationality）作为目标。从而分别产生了 4 种不同的方法论，即图灵测试法、认知建模法、思考规则法，以及理性 Agent 法等。相应地也就有了 4 种不同类型的教科书。这部书把 AI 定义为：设计和建造智能的理性（rational）Agent，并把 Agent 行为的合理性作为评定智能的标准，而不注重这些行为的认知和心理学依据。显见，它是一部注重形式体系并面向工程的书。从方法论看，属于第四类。这也是目前大部分 AI 教科书的编写方式，比较适合于信息科技和其它工程领域的学生和技术人员阅读。

（3）作者十分重视各个部分内容的系统性和相互关联性，采取“突出重点，以点带面”的编写方法，一定程度上避免了 AI 书籍通常存在的内容零散和肤浅的毛病。比如，在知识与推理部分，作者重点分析了比较成熟的一阶逻辑的表示和推理，对它的表达能力和存在的局限性均作了详细的介绍。在此基础上，简单地讨论了其它相关的表示方法，如产生式系统、框架表示和语义网络等。在不确定性和推理部分，用主要篇幅详细分析了概率表示和推理及其相关的信度网络（Belief Networks），而对于默认推理（Default reasoning）、Dempster-Shafer 理论、模糊逻辑等其它可能的方法只作了简单的介绍。从而使整体内容保持严谨、充实与深入，但又不失完整性。

（4）该书对 AI 的重要内容都作了详尽、全面的介绍，而且尽可能采用最新的研究成果。比如，不少书籍在介绍规划（Planning）和学习（Learning）问题时内容比较浅显，往往回避一些关键性的问题（可能由于当时对这些问题的研究还不够深入）。该书在介绍以上两个问题时都尽量引用最新的研究成果，在规划问题上，深入分析了规划与推理的关系，分析了不同规划方法优劣。并根据当前的研究状况，对规划中需要解决的几个难题，如画面问题（Frame Problem，有的译为“框架问题”），条件问题（Qualification Problem）和分支问题（Ramification Problem）等，都给出比较满意的解答。近年来，机器学习研究取得了很大的进展，书中也有较全面的介绍，除决策树以外，还包括了神经网络学习、再励学习（Reinforcement Learning）

等内容。同时讨论了学习中若干基本问题，如学习复杂性、学习精度与泛化（预测）能力等。不过遗憾的是，可能由于篇幅及时间的关系，书中未能包括 Vapnik 等人在统计学习理论上近期所取得的研究成果。这些成果对于理解机器学习的本质会有很大的帮助，希望读者在学习这部分时，可以参考一些最新的文献资料。

（5）理论与实际并重。作者一方面尽量运用数学（或形式化）的语言，对各个领域的原理和方法进行表述，力图让它们建立在严格的理论基础之上。同时，列举了大量的应用实例，这些实例不只限于积木世界和玩具世界，也有不少现实世界的应用。对于那些想把 AI 技术应用于自己所从事的领域的读者来说，无疑会有很大的启发和鼓舞作用。该书每章的后面附有相当数量的练习题和该领域发展历史和介绍，这对于广大读者也很有益处。

（6）虽然这是一部有关智能机器设计的理论基础和方法的书，但作者在最后一章还提出一些 AI 中十分敏感并富有争议的哲学问题：如机器会思考？人工智能能否实现？等等。并讨论了 AI 的现状和未来。这些内容也是广大读者所感兴趣的。

本书可以作为信息领域及相关领域的高等院校本科生和研究生的教科书或教学参考书，也可以作为相关领域的科研与工程技术人员的参考书。

中国科学院院士 清华大学教授

Preface

There are many textbooks that offer an introduction to artificial intelligence (AI). This text has five principal features that together distinguish it from other texts.

1. *Unified presentation of the field.*
   Some texts are organized from a historical perspective, describing each of the major problems and solutions that have been uncovered in 40 years of AI research. Although there is value to this perspective, the result is to give the impression of a dozen or so barely related subfields, each with its own techniques and problems. We have chosen to present AI as a unified field, working on a common problem in various guises. This has entailed some reinterpretation of past research, showing how it fits within a common framework and how it relates to other work that was historically separate. It has also led us to include material not normally covered in AI texts.
2. *Intelligent agent design.*
   The unifying theme of the book is the concept of an *intelligent agent.* In this view, the problem of AI is to describe and build agents that receive percepts from the environment and perform actions. Each such agent is implemented by a function that maps percepts to actions, and we cover different ways to represent these functions, such as production systems, reactive agents, logical planners, neural networks, and decision-theoretic systems. We explain the role of learning as extending the reach of the designer into unknown environments, and show how it constrains agent design, favoring explicit knowledge representation and reasoning. We treat robotics and vision not as independently defined problems, but as occurring in the service of goal achievement. We stress the importance of the task environment characteristics in determining the appropriate agent design.
3. *Comprehensive and up-to-date coverage.*
   We cover areas that are sometimes underemphasized, including reasoning under uncertainty, learning, neural networks, natural language, vision, robotics, and philosophical foundations. We cover many of the more recent ideas in the field, including simulated annealing, memory-bounded search, global ontologies, dynamic and adaptive probabilistic (Bayesian) networks, computational learning theory, and reinforcement learning. We also provide extensive notes and references on the historical sources and current literature for the main ideas in each chapter.
4. *Equal emphasis on theory and practice.*
   Theory and practice are given equal emphasis. All material is grounded in first principles with rigorous theoretical analysis where appropriate, but the point of the theory is to get the concepts across and explain how they are used in actual, fielded systems. The reader of this book will come away with an appreciation for the basic concepts and mathematical methods of AI, and also with an idea of what can and cannot be done with today's technology, at what cost, and using what techniques.
5. *Understanding through implementation.*
   The principles of intelligent agent design are clarified by using them to actually build agents. Chapter 2 provides an overview of agent design, including a basic agent and environment

project. Subsequent chapters include programming exercises that ask the student to add capabilities to the agent, making it behave more and more interestingly and (we hope) intelligently. Algorithms are presented at three levels of detail: prose descriptions and pseudo-code in the text, and complete Common Lisp programs available on the Internet or on floppy disk. All the agent programs are interoperable and work in a uniform framework for simulated environments.

This book is primarily intended for use in an undergraduate course or course sequence. It can also be used in a graduate-level course (perhaps with the addition of some of the primary sources suggested in the bibliographical notes). Because of its comprehensive coverage and the large number of detailed algorithms, it is useful as a primary reference volume for AI graduate students and professionals wishing to branch out beyond their own subfield. We also hope that AI researchers could benefit from thinking about the unifying approach we advocate.

The only prerequisite is familiarity with basic concepts of computer science (algorithms, data structures, complexity) at a sophomore level. Freshman calculus is useful for understanding neural networks and adaptive probabilistic networks in detail. Some experience with nonnumeric programming is desirable, but can be picked up in a few weeks study. We provide implementations of all algorithms in Common Lisp (see Appendix B), but other languages such as Scheme, Prolog, Smalltalk, C++, or ML could be used instead.

Overview of the book

The book is divided into eight parts. Part I, "Artificial Intelligence," sets the stage for all the others, and offers a view of the AI enterprise based around the idea of intelligent agents—systems that can decide what to do and do it. Part II, "Problem Solving," concentrates on methods for deciding what to do when one needs to think ahead several steps, for example in navigating across country or playing chess. Part III, "Knowledge and Reasoning," discusses ways to represent knowledge about the world—how it works, what it is currently like, what one's actions might do—and how to reason logically with that knowledge. Part IV, "Acting Logically," then discusses how to use these reasoning methods to decide what to do, particularly by constructing *plans*. Part V, "Uncertain Knowledge and Reasoning," is analogous to Parts III and IV, but it concentrates on reasoning and decision-making in the presence of *uncertainty* about the world, as might be faced, for example, by a system for medical diagnosis and treatment.

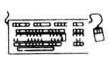
Together, Parts II to V describe that part of the intelligent agent responsible for reaching decisions. Part VI, "Learning," describes methods for generating the knowledge required by these decision-making components; it also introduces a new kind of component, the *neural network*, and its associated learning procedures. Part VII, "Communicating, Perceiving, and Acting," describes ways in which an intelligent agent can perceive its environment so as to know what is going on, whether by vision, touch, hearing, or understanding language; and ways in which it can turn its plans into real actions, either as robot motion or as natural language utterances. Finally, Part VIII, "Conclusions," analyses the past and future of AI, and provides some light amusement by discussing what AI really is and why it has already succeeded to some degree, and airing the views of those philosophers who believe that AI can never succeed at all.

Using this book

This is a big book; covering *all* the chapters and the projects would take two semesters. You will notice that the book is divided into 27 chapters, which makes it easy to select the appropriate material for any chosen course of study. Each chapter can be covered in approximately one week. Some reasonable choices for a variety of quarter and semester courses are as follows:

- *One-quarter general introductory course*:
  Chapters 1, 2, 3, 6, 7, 9, 11, 14, 15, 18, 22.
- *One-semester general introductory course*:
  Chapters 1, 2, 3, 4, 6, 7, 9, 11, 13, 14, 15, 18, 19, 22, 24, 26, 27.
- *One-quarter course with concentration on search and planning*:
  Chapters 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13.
- *One-quarter course with concentration on reasoning and expert systems*:
  Chapters 1, 2, 3, 6, 7, 8, 9, 10, 11, 14, 15, 16.
- *One-quarter course with concentration on natural language*:
  Chapters 1, 2, 3, 6, 7, 8, 9, 14, 15, 22, 23, 26, 27.
- *One-semester course with concentration on learning and neural networks*:
  Chapters 1, 2, 3, 4, 6, 7, 9, 14, 15, 16, 17, 18, 19, 20, 21.
- *One-semester course with concentration on vision and robotics*:
  Chapters 1, 2, 3, 4, 6, 7, 11, 13, 14, 15, 16, 17, 24, 25, 20.

These sequences could be used for both undergraduate and graduate courses. The relevant parts of the book could also be used to provide the first phase of graduate specialty courses. For example, Part VI could be used in conjunction with readings from the literature in a course on machine learning.



We have decided *not* to designate certain sections as "optional" or certain exercises as "difficult," as individual tastes and backgrounds vary widely. Exercises requiring significant programming are marked with a keyboard icon, and those requiring some investigation of the literature are marked with a book icon. Altogether, over 300 exercises are included. Some of them are large enough to be considered term projects. Many of the exercises can best be solved by taking advantage of the code repository, which is described in Appendix B. Throughout the book, important points are marked with a *pointing icon*.

If you have any comments on the book, we'd like to hear from you. Appendix B includes information on how to contact us.

Acknowledgements

Jitendra Malik wrote most of Chapter 24 (Vision) and John Canny wrote most of Chapter 25 (Robotics). Doug Edwards researched the Historical Notes sections for all chapters and wrote much of them. Tim Huang helped with formatting of the diagrams and algorithms. Maryann Simmons prepared the 3-D model from which the cover illustration was produced, and Lisa Marie Sardegna did the postprocessing for the final image. Alan Apt, Mona Pompili, and Sondra Chavez at Prentice Hall tried their best to keep us on schedule and made many helpful suggestions on design and content.

Stuart would like to thank his parents, brother, and sister for their encouragement and their patience at his extended absence. He hopes to be home for Christmas. He would also like to thank Loy Sheflott for her patience and support. He hopes to be home some time tomorrow afternoon. His intellectual debt to his Ph.D. advisor, Michael Genesereth, is evident throughout the book. RUGS (Russell's Unusual Group of Students) have been unusually helpful.

Peter would like to thank his parents (Torsten and Gerda) for getting him started, his advisor (Bob Wilensky), supervisors (Bill Woods and Bob Sproull) and employer (Sun Microsystems) for supporting his work in AI, and his wife (Kris) and friends for encouraging and tolerating him through the long hours of writing.

Before publication, drafts of this book were used in 26 courses by about 1000 students. Both of us deeply appreciate the many comments of these students and instructors (and other reviewers). We can't thank them all individually, but we would like to acknowledge the especially helpful comments of these people:

Tony Barrett, Howard Beck, John Binder, Larry Bookman, Chris Brown, Lauren Burka, Murray Campbell, Anil Chakravarthy, Roberto Cipolla, Doug Edwards, Kutluhan Erol, Jeffrey Forbes, John Fosler, Bob Futrelle, Sabine Glesner, Barbara Grosz, Steve Hanks, Othar Hansson, Jim Hendler, Tim Huang, Seth Hutchinson, Dan Jurafsky, Leslie Pack Kaelbling, Keiji Kanazawa, Surekha Kasibhatla, Simon Kasif, Daphne Koller, Rich Korf, James Kurien, John Lazzaro, Jason Leatherman, Jon LeBlanc, Jim Martin, Andy Mayer, Steve Minton, Leora Morgenstern, Ron Musick, Stuart Nelson, Steve Omohundro, Ron Parr, Tony Passera, Michael Pazzani, Ira Pohl, Martha Pollack, Bruce Porter, Malcolm Pradhan, Lorraine Prior, Greg Provan, Philip Resnik, Richard Scherl, Daniel Sleator, Robert Sproull, Lynn Stein, Devika Subramanian, Rich Sutton, Jonathan Tash, Austin Tate, Mark Torrance, Randall Upham, Jim Waldo, Bonnie Webber, Michael Wellman, Dan Weld, Richard Yen, Shlomo Zilberstein.

Summary of Contents

Contents