INTEGRATING RULES and CONNECTIONISM for ROBUST COMMONSENSE REASONING

Ron Sun

Sixth-Generation Computer Technology Series
Branko Souček, Series Editor

Integrating Rules and Connectionism for Robust Commonsense Reasoning

RON SUN
Department of Computer Science
The University of Alabama
Tuscaloosa, Alabama

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York - Chichester - Brisbane - Toronto - Singapore

This text is printed on acid-free paper.

Library of Congress Cataloging in Publication Data:

Sun, Ron, 1960–
Integrating rules and connectionism for robust commonsense reasoning/ Ron Sun.
p. cm. — (Sixth-generation computer technology series)
Includes index.
ISBN 0-471-59324-9 (alk. paper)
1. Case-based reasoning. 2. Cognition. 3. Neural computers.
I. Title. II. Series.
Q338.8.S85 1994
006.3—dc20 93-19460

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

FOREWORD

When Ron Sun first proposed his ideas for achieving commonsense reasoning, I was skeptical. This must have been around 1988 at Brandeis University, where he was my Ph.D. advisee. Ron envisioned a hybrid system—CONSYDERR—with a rule-based portion and a connectionist portion, joined by links, solving the same problems in parallel, and influencing each other during different phases of an overall system schedule. Despite my general belief in the fundamental correctness of Minsky's society of mind ideas concerning cognitive architecture in people, it was easy to be more critical of Ron's specific architecture than of Minsky's compendium of design ideas and observations. Why just two levels? Why these particular two levels?

Three years later, while I still retained some level of skepticism, Ron showed that his system could handle every example I threw at him. This gave me sufficient confidence in his dissertation. Subsequently, he was selected for the David Marr Award at the Cognitive Science Conference in 1991, which further confirmed the importance of this work. More recently, I have encountered some ideas and problems that cause me to think that Ron is on to something even bigger. I have come to see that his work is closely related to some deep issues in the evolutionary "design" of the brain and in software engineering. Let me explain.

Evolution in the brain has taken place by accretion of new layers on top of relatively unchanged old structures, rather than by mutation of these old structures. Thus the brain stem and midbrain are relatively unchanged from those of other primates; our big differences are in the growth of the cortex on top of (or around) these older layers. Why should evolution work this way? Presumably, once various functions are worked out, it is maladaptive to modify them in almost any fashion. However, it apparently is beneficial to add new structures (e.g., the cortex) around the old ones (e.g., the midbrain) that may trap conditions that would not be understood by (or that might be misunderstood by) the older structures. Such structures can clearly be captured in a two-level model.

What does this have to do with software engineering? With the discoveries of neuroscientists now flooding the scientific literature as well as the daily papers, I believe there will be great opportunities for designing computer systems that operate on principles quite different from those used in today's general purpose computers. However, one cannot imagine scrapping our software legacy of operating systems, compilers, and applications—trusted software that operates in many cases continuously with 99.9% uptime. Many systems that are in constant use were written 20 or more years ago. Many of the people who wrote them have retired, and no one can understand 20-year-old code in any case. It is rumored that many mainframe applications will cease to function on January 1, 2000. On that date, ancient COBOL programs will subtract 99 from 00 (i.e., the two-digit encoding for the years 2000 and 1999), generating an error for each such operation. Another special kind of problem arises because these applications were written when memory was very precious, and so used many tricks—such as combining several operations in one instruction—which make the old code virtually undecipherable. Yet we will certainly want to exploit novel architectures, e.g., massively parallel, neurally inspired systems. But it's hard to imagine how we could ever translate legacy software to run on such systems.

What can we do to break this impasse? One possibility is to build hybrid systems, with traditional general-purpose computers surrounded by systems/devices of new design, with sophisticated communication channels between them. These surrounding machines can pass requirements through to old devices when appropriate, but at the same time can trap novel or modified operations that cannot be dealt with by the old chips. This is highly reminiscent of Ron's design for CONSYDERR! It shows the universality of two-level structuring.

With enough examples from radically different realms, I see that important principles are coming out of this work. I therefore recommend Ron's book with great enthusiasm.

DAVID L. WALTZ

Princeton, New Jersey

PREFACE

One of the more difficult problems for artificial intelligence research is the problem of modeling commonsense reasoning. The use of traditional models to capture the flexible and robust nature of commonsense reasoning presents great difficulties. In this book I attempt to tackle this problem by adopting innovative approaches. In a nutshell, I am concerned with understanding and modeling commonsense reasoning with a combination of rules and similarities, under a connectionist rubric. I survey the areas of reasoning, connectionist models, inheritance, causality, rule-based systems, and similarity-based reasoning, and introduce a new framework and a novel connectionist architecture for modeling commonsense reasoning that synthesizes some of these areas. Along with this framework, I describe a set of interrelated new ideas regarding modeling commonsense reasoning that are very relevant to current artificial intelligence and cognitive science research and the ongoing methodological debate.

To make these issues clearer, I will describe very briefly the main thrust of this work. First, some reasoning data and examples are analyzed to gain an insight into the matter. This analysis establishes a framework for modeling such data based on the notions of rules and similarities, and then, through detailed thought experiments and derivations, the framework is translated into a connectionist architecture. The architecture is capable of carrying out both rule-based reasoning and similarity-based reasoning naturally, for similarity-based reasoning is inherent in connectionist models and rule-based reasoning is integrated into connectionist networks through encoding rules in these networks.

Technical details aside, an important conclusion from the research described in this book is that the synergy resulting from the interaction of the two different types of representation and processing (such as rule-based reasoning and similarity-based reasoning) can and should be utilized. This synergy enables the handling of a large number of difficult issues in commonsense reasoning, within *one* integrated framework. The results presented in the book suggest that connectionist models incorporating rule-based reasoning can be effective and efficient models of reasoning.

Based on this architecture, a number of technical and theoretical issues are addressed in the book. For example, to understand rule encoding in connectionist architectures, a formal analysis of the rule encoding scheme is performed. Unlike previous work, the analysis shows that it handles a superset of some important logics. To further improve the rule-based reasoning capability of this connectionist architecture, variables are added, and a solution to the connectionist *variable binding* problem is proposed which has a number of advantages over existing solutions. I also explore the notion of *causality* in the connectionist model and show that commonsense causal knowledge can be well represented in connectionist frameworks in general and in this architecture in particular. Several other aspects of the architecture are also discussed in the book to demonstrate how connectionist models can supplement, enhance, and integrate symbolic rule-based reasoning in tackling commonsense reasoning.

This book is aimed at readers who are interested in the broad subject of modeling and understanding cognitive processes, especially commonsense reasoning, who are interested in overcoming the limitations of currently available computational machinery and striving for a new generation of commonsensical intelligent systems, and who are interested in a multidisciplinary attack on AI problems based on the confluence of many different intellectual sources. Specifically, it should be useful for researchers and students in artificial intelligence, cognitive science, neural networks, and other related areas. It should also be suitable for interested laypersons who wish to be informed of the new research being carried out.

HISTORY OF THE WORK

This book is adapted from my Ph.D. dissertation, completed at Brandeis University. The research it describes was started during the summer of 1988 and carried out during my three years in the Michtom School of Computer Science at Brandeis University. Some of the ideas in this book, however, can be traced far back in time, to early 1984. During that period, while I was reading psychology and AI literatures, some ideas about human (and machine) reasoning started to come to my mind, I did not get a chance to pursue these ideas any further until the summer of 1988 when I finally had the opportunity to do so. I think I made the right decision to go to Brandeis University. But it was never easy, I guess no matter where I was, to go through the process of forming a new idea, presenting it convincingly, and gaining support for it. The most important thing I have learned during my graduate school years is how to narrow down a goal, form a reasonable agenda, and accomplish the agenda. By the late summer of 1990, I had accomplished a reasonable number of items on my agenda and started the actual writing of the dissertation. The writing was completed in the summer of 1991.

After completing the dissertation, I started to write journal papers based on the work reported in the dissertation. I received a large number of requests for copies of my dissertation (given the size of the work, I could not possibly fulfill all these requests). Several people who read my papers and/or the dissertation suggested to me the idea of publishing it in book form. I signed the book contract with John Wiley & Sons, Inc., in the summer of 1992, eight years after I first had the idea and four years after I started actually doing the research.

The manuscript was thoroughly revised and augmented since then. The revision took another six months. The book is the final outcome of this long process.

ACKNOWLEDGMENTS

I wish to take this opportunity to acknowledge the following people. First of all, during my years at Brandeis University, many people have helped me in pursuing this research: I want to thank Dave Waltz for his constant guidance and for his careful reading of numerous rough drafts; I want to thank James Pustejovsky for spending many hours with me discussing the research and for his extremely helpful comments on the manuscript; I want to thank Tim Hickey for many very useful discussions that have helped me to clarify formal aspects of my work; Pattie Maes also provided valuable feedback. Jerry Samet introduced me to the philosophy of mind. Larry Bookman wrote detailed comments on my papers, and the frequent discussions with him were thought-provoking.

During the publication process that transformed the original dissertation into this book, I have received comments and feedback from a number of colleagues, to whom I owe my gratitude. Dan Levine provided thorough, thoughtful, and inspiring comments and detailed suggestions for revisions, as well as reference materials. Branko Souček provided positive and encouraging comments that led to the publication of the book. Steve Sloman commented extensively on my work, thus helping me to greatly clarify my thinking about the issues. Thanks also go to several anonymous reviewers for various suggestions.

I have benefited from interactions with many others (names are listed in random order): Jim Hendler, Lokendra Shastri, Tim Finin, Jude Shavlik, Ryszard Michalski, Gomer Thomas, George Mou, Rick Alterman, Allan

Collins, Helen Gigley, Shashi Shekhar, John Barnden, Michael Dyer, Dave Touretzky, Jordan Pollack, and Michael Arbib. Others have undoubtedly been omitted, due to my less than perfect memory.

I wish to thank the Wiley editors for their professional help.

I acknowledge the financial support that I received during the years I carried out the work that led to this book. My three-year stay at Brandeis was supported by fellowships, scholarships, and assistantships provided by the university and indirectly from various funding agencies. My work at the University of Alabama (during the revision of the manuscript) was supported by the university and its College of Engineering. The David Marr Award from the Cognitive Science Society is also gratefully acknowledged.

CONTENTS

ILLUSTRATIONS

CHAPTER 1

Introduction

The complexity of philosophy is not in its matter, but in our tangled understanding.
—L. Wittgenstein, PHILOSOPHICAL REMARKS

1.1 OVERVIEW

The dream of building truly intelligent machines that are as capable as human beings and that can interact with people in a natural and meaningful way has had a long history, as demonstrated by the work of Leibniz, Boole, Turing, and many others (see, e.g., Turing 1950). It predates artificial intelligence, a field that investigates computational ways of capturing intelligence (formed in 1956 with the Dartmouth conference, as commonly recognized). Artificial intelligence (AI) builds on this age-old dream: it is in fact the first serious and massive effort at understanding and capturing intelligence computationally, and it has been inspired by the invention and the gradual popularization of digital computers; for the success of simple logics (carried out by hardware switching elements) in formulating arithmetic computations in digital computers led to the hope that they could also capture more "intelligent" processes, such as problem solving and planning, up to the level of complete human intelligence.

In nearly four decades following the inception of AI, it has had some successes in formulating many processes that were thought to require high-level "intelligence," such as chess, medical diagnosis, and symbolic mathematics (Nilsson 1980, Buchanan and Shoftliffe 1984). However, none of the successful AI systems have real "common sense," the ability of human beings to understand simple, common things and to get to the heart of a problem without enumerating all the rules and following them one by one.

The problem of common sense and commonsense reasoning has become one of the central issues in artificial intelligence: How can we capture commonsense reasoning *computationally*, both effectively and efficiently in terms of computational resources? The question can be broken down into a number of subquestions based on our current understanding of this issue and all the available scientific theories:

- How can we conceptually characterize commonsense reasoning in a simple and uniform way? What kinds of formalisms and models are suitable?
- How can real AI systems capture common sense and commonsense reasoning?
- What are the problems and difficulties with existing approaches?
- Can connectionist models, which are relatively new on the horizon, deal with commonsense reasoning?
- More important, can connectionist models do better than existing approaches in accounting for commonsense reasoning?

This work is intended to examine such questions. What this work shows is that connectionist models of reasoning can implement some traditional "symbolic" AI paradigms, but they are not mere "implementations" of their symbolic counterparts; rather, they can be better computational models of commonsense reasoning than the traditional ones. They can be better because they take into consideration the approximate, evidential, and flexible nature of commonsense reasoning, and they also account for the spontaneity and parallelism in commonsense reasoning, which are very important but often neglected.

This chapter provides an introduction that motivates the work described in subsequent chapters, as well as to provide a summary of what this book is all about. In the rest of this chapter, I first briefly review various approaches to commonsense reasoning, and then I introduce the problem of modeling common patterns in commonsense reasoning and explain the difficulties involved. Then I outline a solution to the problem. The rationales for this research are enumerated next, followed by some clarifications. Some guides for readers are provided at the end of the chapter.

1.2 COMMONSENSE REASONING

1.2.1 The Problem of Commonsense Reasoning

Reasoning is a fundamental aspect of human cognition. It is involved in all kinds of cognitive processes, ranging from language understanding to deci-