LEARNING FROM GOOD AND BAD DATA

THE KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCES

KNOWLEDGE REPRESENTATION, LEARNING AND EXPERT SYSTEMS

Consulting Editor
Tom Mitchell
Carnegie Mellon University

Other books in the series:

Universal Subgoaling and Chunking of Goal Hierarchies, J. Laird, P. Rosenbloom, A. Newell. ISBN: 0-89838-213-0.

Machine Learning: A Guide to Current Research, T. Mitchell, J. Carbonell, R. Michalski. ISBN: 0-89838-214-9.

Machine Learning of Inductive Bias, P. Utgoff. ISBN: 0-89838-223-8.

A Connectionist Machine for Genetic Hillclimbing, D. H. Ackley. ISBN: 0-89838-236-x.

LEARNING FROM GOOD AND BAD DATA

by

Philip D. Laird
NASA Ames Research Center

KLUWER ACADEMIC PUBLISHERS
Boston/Dordrecht/Lancaster

Distributors for North America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061, USA

Distributors for the UK and Ireland:
Kluwer Academic Publishers
MTP Press Limited
Falcon House, Queen Square
Lancaster LA1 1RN, UNITED KINGDOM

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data

Laird, Philip D.
Learning from good and bad data / Philip D. Laird.
p. cm. — (The Kluwer international series in engineering and computer science ; 47. Knowledge representation, learning, and expert systems)
Bibliography: p.
Includes index.

ISBN-13: 978-1-4612-8951-7 e-ISBN-13: 978-1-4613-1685-5

DOI: 10.1007/978-1-4613-1685-5

1. Machine learning. 2. Artificial intelligence. 3. Computers. 4. System identification. I. Title. II. Series: Kluwer international series in engineering and computer science ; SECS 47. III. Series: Kluwer international series in engineering and computer science. Knowledge representation, learning, and expert systems.
Q325.L35 1988 87-35103
006.3'1—dc19 CIP

Softcover reprint of the hardcover 1st edition 1988

Contents

List of Figures

PREFACE

This monograph is a contribution to the study of the ***identification problem***: the problem of identifying an item from a known class using positive and negative examples. This problem is considered to be an important component of the process of inductive learning, and as such has been studied extensively. In the overview we shall explain the objectives of this work and its place in the overall fabric of learning research.

Context. Learning occurs in many forms; the only form we are treating here is inductive learning, roughly characterized as the process of forming general concepts from specific examples. Computer Science has found three basic approaches to this problem:

- Select a specific learning task, possibly part of a larger task, and construct a computer program to solve that task.
- Study cognitive models of learning in humans and extrapolate from them general principles to explain learning behavior. Then construct machine programs to test and illustrate these models.

- Formulate a mathematical theory to capture key features of the induction process.

This work belongs to the third category.

The various studies of learning utilize training examples (data) in different ways. The three principal ones are:

- Similarity-based (or empirical) learning, in which a collection of examples is used to select an explanation from a class of possible rules.
- Explanation-based learning, in which each example is analyzed individually within the context of a base of knowledge about the domain, the learning goals, and other non-information-theoretic properties. The result of the analysis is then generalized, thereby enlarging the knowledge base.
- Network models of learning, in which large numbers of similar elementary units are interconnected, usually in a fixed topology. The network parameters can be adjusted as examples are presented, until the network as a whole exhibits the desired behavior.

The work described here pertains only to the first approach.

A significant collection of similarity-based theoretical results exists for inferring classes of recursive functions from examples of their values at sample points of the domain. (See, for example, [14], [19], [44].) I shall not adopt this approach. Instead, I prefer to focus on less expressive, more specific domains whose concrete properties can be utilized in constructing the algorithms.

Scope of the Field. The overall goal of this, and related, research is to devise and study good mathematical models of the inductive-inference process. Although many papers with this purpose have been written, the "right" model (by general consensus) has not emerged. What should a good model accomplish? As a minimum, the model should be descriptive of many of the domains and techniques used in artificial intelligence (AI) experiments and applications. And it should be specific enough to permit us to devise algorithms and to reason about their complexity. That this has not yet been done may be illustrated by the fact that there is no general agreement about the requirements of an inductive inference domain, what exactly is meant by a "training example", or what the most appropriate methods are for presenting information.

Once one has proposed such a model, he or she must then determine what the model has to say about previous research, both experimental and mathematical. Inevitably the model will succeed in capturing certain aspects but not others. We often learn as much from the deficiencies of a model as from its successes, for the fact that a particular phenomenon of induction cannot be implemented within the current model implies that the phenomenon entails new ideas and is probably worthy of independent research.

Justification. Judging from the rapid growth of interest in the field, there is little disagreement that inductive learning is an important problem for computer science. Two of the applications most often envisioned for a successful theory are the automatic construction and refinement of large knowledge bases for expert systems and the recognition and classification of data patterns for robotics control software. Apart from these, there is the significant scientific problem of understanding a fundamental and widespread phenomenon such as learning. Even though we cannot even define precisely what learning is, we have a strong sense that there are profound and unifying principles underlying the diversity of behaviors that apparently exhibit some form of learning. To give an analogy, the study of the physics of gases in the mid-nineteenth century proceeded on the basis of the assumption that the many observed laws of gases could be explained by suitably applying the fundamental laws of mechanics. The outgrowth of this idea was the theory of statistical mechanics, which was both mathematically satisfying and useful for engineering applications and further scientific research.

But why a mathematical theory? Are we yet to the point where we understand the problem of inductive learning well enough to formulate it mathematically? To date, theoretical research in inductive inference has played a rather minor part in the design and implementation of computer programs, and it is unlikely that the results in this report will alter this state of affairs.

Nevertheless we need the theory, for these reasons:

- Theory, whether mathematical or not, serves to distill from the mass of experimental details that core of ideas that apply to a more general range of problem settings. A recent critical review of the research noted that "a significant problem in current research on inductive learning is that each research goup is using a different notation and terminology. This not only makes the exchange of research results difficult, but it also makes it hard for new researchers to enter the field." ([22]). Noteworthy cases where theory *has* identified general concepts include the Version Space theory of [46], the Model Inference System of Shapiro ([65]), and the work on approximation by sampling ([71], [15]).

- An essential aspect of theoretical modeling is ***simplification***. In exchange for the detail the theory leaves out, one gains the perspective of which ideas should be given priority, the ability to reason more clearly about the problem, and the potential to unify phenomena that would otherwise appear unrelated. An instance of this process at work is the research of Shapiro ([66]) in which the inductive inference of logical theories, the identification of functions, and the systematic debugging of programs are all shown to be closely related.
- Even when the theory contributes nothing to practice, it may strongly affect the way we conceptualize the problem. No programmer constructing or coding an algorithm refers explicitly to the theory of Turing machines (or any other basic model of effective computation). Yet the concept of a finite control with an infinite store dominates the way we think about sequential computation.

Organization. The results that follow are in two parts. Part One develops a model for the identification problem that captures the strategy of generalization and specialization of hypotheses based on counterexamples. This approach to the identification problem has been the basis of much research over the years.

Chapter One introduces our formal model of the identification problem, when rules are to be learned in the limit from an arbitrary but thorough teacher. The model is general enough to apply directly to a wide class of commonly studied domains.

Chapter Two defines the relations that are used in generalizing and specializing. For unity we call both the generalizing and the specializing relations *refinement relations*, since they enjoy the same algebraic properties. We also present general algorithms for the identification problem using refinements.

Chapter Three considers the problem of designing refinements that are less general but more useful. To do so, we consider domains with specific, but common, properties and devise relations that take advantage of these properties. In addition we show how to construct refinement relations for refinements themselves, and argue that the search for an appropriate inductive bias can be understood as a process of refinement.

From this theory, we gain the following:

- A clear definition of what constitutes a domain for an identification problem, what we mean by "examples", and a simple characterization of the operators that are useful for generalization and specialization.
- Abstraction of many of the common aspects of different identification algorithms.
- Recognition of some basic limitations of this approach, particularly the fixed representation language and the lack of meaningful complexity measures for comparing algorithms.

In Part Two we adopt a different model of the identification problem, one in which examples are drawn independently from an unknown distribution. Chapter Four defines the model and two criteria for identication: stochastic identification in the limit, and *pac*-identification as defined by Valiant ([71]). An advantage of this model is that we can measure the complexity of identification over many domains. Another is that the effects of random noise in the examples can be studied: this is the subject of Chapter Five. From the analysis we reach a clear understanding of how the strategy of identification from noisy examples generalizes the corresponding strategy for reliable examples. Thus along with useful mathematical results, such as bounds for the number of training examples required, we obtain a general conceptual principle for the design of noise-tolerant identification algorithms.

About the presentation. In order that the contents of this book be accessible to more than just those in the fields of machine learning and theoretical computer science, I have chosen to present the mathematical results with rather more motivational and expository remarks than is currently fashionable. My hope is that readers will be able follow the main lines of the development without reading the mathematical details.

Acknowledgments This research has benefitted immeasurably from the kind assistance of many people. It is my great pleasure to acknowledge their contributions.

Dana Angluin served as my adviser during my four years at Yale. Lured by the insight and elegance of her work and by her reputation as a fine teacher, I came to Yale with the hope of studying with her. Despite the competing demands of family, students, and research, she always found time to listen to half-formed ideas, wade through poorly written drafts, and offer numerous suggestions (technical and otherwise).

Dana has also contributed directly to the content of this monograph. In particular, much of Chapter 5 is based on our joint research ([6]). Section 5.6 is mostly her ideas, but because of its relevance she has allowed me to include my version of it here. Also, I have used her (more elegant) proof of Theorem 5.9 in preference to my own. The observations of Section 3.4.1 are primarily Dana's as well.

Neil Immerman and Tom Mitchell bravely agreed to serve as readers, each providing valuable input from the perspective of his own field of expertise.

Conversations with Ranan Banerji, Josh Benaloh, Sandeep Bhatt, Peter Dunning, David Haussler, Neil Immerman, Lenny Pitt, Takeshi Shinohara, and many others during the course of my research were very helpful. At NASA-Ames Roger Bartlett gave generously of his time to help me run off this manuscript. And finally, my colleagues and teachers at Yale endured presentations of these results and provided an environment of friendship and stimulation.

I am grateful to the folks at Kluwer Academic for the opportunity to make my ideas available to a wider readership.

In large part this research was funded by the National Science Foundation, under grants MCS-8002447 and IRI-8404226. Without broad-based funding of this type, the fundamental research so essential to our collective progress would not be possible.