

Cognitive Technologies

Managing Editors: D. M. Gabbay J. Siekmann

Editorial Board: A. Bundy J. G. Carbonell
M. Pinkal H. Uszkoreit M. Veloso W. Wahlster
M. J. Wooldridge

Advisory Board:

Luigia Carlucci Aiello
Franz Baader
Wolfgang Bibel
Leonard Bolc
Craig Boutilier
Ron Brachman
Bruce G. Buchanan
Anthony Cohn
Artur d'Avila Garcez
Luis Fariñas del Cerro
Koichi Furukawa
Georg Gottlob
Patrick J. Hayes
James A. Hendler
Anthony Jameson
Nick Jennings
Aravind K. Joshi
Hans Kamp
Martin Kay
Hiroaki Kitano
Robert Kowalski
Sarit Kraus
Maurizio Lenzerini
Hector Levesque
John Lloyd

Alan Mackworth
Mark Maybury
Tom Mitchell
Johanna D. Moore
Stephen H. Muggleton
Bernhard Nebel
Sharon Oviatt
Luis Pereira
Lu Ruqian
Stuart Russell
Erik Sandewall
Luc Steels
Oliviero Stock
Peter Stone
Gerhard Strube
Katia Sycara
Milind Tambe
Hidehiko Tanaka
Sebastian Thrun
Junichi Tsujii
Kurt VanLehn
Andrei Voronkov
Toby Walsh
Bonnie Webber

Pavel Brazdil · Christophe Giraud-Carrier
Carlos Soares · Ricardo Vilalta

Metalearning

Applications to Data Mining

With 53 Figures and 11 Tables

Authors:

Prof. Pavel Brazdil
LIAAD
Universidade do Porto
Fac. Economia
Rua de Ceuta 118-6°
4050-190 Porto, Portugal
pbrazdil@liaad.up.pt

Dr. Christophe Giraud-Carrier
Brigham Young University
Department of Computer Science
Provo, UT 84602, USA
cgc@cs.byu.edu

Dr. Carlos Soares
LIAAD
Universidade do Porto
Fac. Economia
Rua de Ceuta 118-6°
4050-190 Porto, Portugal
csoares@fep.up.pt

Dr. Ricardo Vilalta
University of Houston
Department of Computer Science
501 PGH Building
Houston, TX 77204-3010, USA
vilalta@cs.uh.edu

Managing Editors:

Prof. Dov M. Gabbay
Augustus De Morgan Professor of Logic
Department of Computer Science
King's College London
Strand, London WC2R 2LS, UK

Prof. Dr. Jörg Siekmann
Forschungsbereich Deduktions- und
Multiagentensysteme, DFKI
Stuhlsatzenweg 3, Geb. 43
66123 Saarbrücken, Germany

ISBN: 978-3-540-73262-4 e-ISBN: 978-3-540-73263-1
DOI: 10.1007/978-3-540-73263-1

Cognitive Technologies ISSN: 1611-2482

Library of Congress Control Number: 2008937821

ACM Computing Classification (1998): I.2.6, H.2.8

© Springer-Verlag Berlin Heidelberg 2009

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: KünkelLopka, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Dedication

Pavel

to my wife and lifelong companion, Fátima.

Christophe

to my wife and children.

Carlos

to Manela, Quica and Manel.

Ricardo

to my parents.

Preface

The continuous growth of successful applications in machine learning and data mining has led to an apparent view of real progress towards the understanding of the nature and mechanisms of learning machines. From the automated classification of millions of luminous objects stored in astronomical images, to the complex analysis of long sequences of genes in biomedical data sets, machine learning has positioned itself as an indispensable tool for data analysis, pattern recognition, and scientific discovery. This apparent progress in the search for accurate predictive models relies on the design of learning algorithms exhibiting novel functionalities. The history of machine learning shows a research community devoted to the study and improvement of a highly diverse set of learning algorithms such as nearest neighbors, Bayesian classifiers, decision trees, neural networks, and support vector machines (to name just a few). While the design of new learning algorithms is certainly important in advancing our ability to finding accurate data models, so is the understanding of the relation between data set characteristics and the particular mechanisms embedded in the learning algorithm. Rather than testing multiple algorithms to assess which one would perform satisfactorily on a certain data set, the end user needs guidelines pointing to the best learning strategy for the particular problem at hand. Researchers and practitioners in machine learning have a clear need to answer the following question: what works well where? There is a strong need to characterize both data distributions and learning mechanisms to construct a theory of learning behavior. Moreover, we advocate the development of a new generation of learning algorithms that are capable of profound adaptations in their behavior to the input data. This may include changes to the model language itself.

Despite different interpretations of the term *metalearning*, in this book we pursue the goal of finding principled methods that can make learning algorithms adaptive to the characteristics of the data. This can be achieved in many ways as long as there is some form of feedback relating learning performance with data distributions. Thus one can think of the problem of algorithm selection (or ranking), or algorithm combination, as frameworks

that exploit past performance to guide the selection of a final model. The ultimate end is to design learning algorithms that adapt to the problem at hand, rather than invoking the same fixed mechanisms independent of the nature of the data under analysis. We explain in this book that a unifying theme in metalearning is that of exploiting experience or metaknowledge to achieve flexible learning systems.

These ideas have brought us together to write a book that summarizes the current state of the art in the field of metalearning. The motivation for such a book can be traced back to the METAL project [166], in which the first three authors were active participants. Our first (electronic) meetings regarding this book took place in the second half of 2005 and have continued over the following three years until the second half of 2008. The project proved challenging in many ways, most particularly in unifying our view concerning the scope of the term metalearning. After long discussions we finally agreed on the definition provided in Chapter 1. Equally challenging was to decide on a list of topics that stand as clearly representative of the field. We hope the reader will find our selection appropriate and sufficiently broad to offer adequate coverage. Finally, it is our hope that this book will serve not only to place many of the ideas now dispersed all over the field into a coherent volume, but also to encourage researchers in machine learning to consider the importance of this fascinating area of study.

Book Organization

The current state of diverse ideas in the field of metalearning is not yet mature enough for a textbook based on a solid conceptual and theoretical framework of learning performance. Given this, we have decided to cover the main topics where there seems to be a clear consensus regarding their relevance and legitimate membership in the field. In the following, we briefly describe the contents of the book acknowledging the contribution of each of the authors. In Chapter 1, all of us worked on introducing the main ideas and concepts that we believe are essential to understanding the field of metalearning. Chapters 2–4 have a more practical flavor, illustrating the important problem of selecting and ranking learning algorithms, with a description of several currently operational applications. In Chapter 2, Soares and Brazdil describe a simple meta-learning system that, given a dataset, provides the user with guidance concerning which learning algorithm to use. The issues that are involved in developing such a system are discussed in more detail by Soares in Chapter 3, including a survey of existing approaches to address them. In Chapter 4, Giraud-Carrier describes a number of systems that incorporate some form of automatic user guidance in the data mining process.

Chapters 5–7, on the other hand, have a more conceptual flavor, covering the combination of classifiers, learning from data streams, and knowledge transfer. Chapter 5, authored by Giraud-Carrier, describes the main concepts behind model combination, including classical techniques such as bagging and

boosting, as well as more advanced techniques such as delegating, arbitrating and meta-decision trees. We invited Gama and Castillo to contribute to Chapter 6; the chapter discusses the dynamics of the learning process and general strategies for reasoning about the evolution of the learning process itself. The main characteristics and new constraints on the design of learning algorithms imposed by large volumes of data that evolve over time are described, including embedding change-detection mechanisms in the learning algorithm and the trade-off between the cost of update and the gain in performance. Chapter 7, authored by Vilalta, covers the important topic of knowledge transfer across tasks; the chapter covers topics such as multitask learning, transfer in kernel methods, transfer in parametric Bayesian methods, theoretical models of learning to learn, and new challenges in transfer learning with examples in robotics. Lastly, Chapter 8, authored by Brazdil, discusses the important role of metalearning in the construction of complex systems through the composition of induced subsystems. It is shown how domain-specific metaknowledge can be used to facilitate this task.

Acknowledgements

We wish to express our gratitude to all those who helped in bringing this project to fruition. We are grateful to the University of Porto and Faculty of Economics and also to the Portuguese funding organization FCT for supporting the R&D laboratory LIAAD (Laboratory of Artificial Intelligence and Decision Support) where a significant part of the work associated with this book was carried out. We also acknowledge support from the Portuguese funding organization FCT for the project ALES II – Adaptive LEarning Systems. This work was also partially supported by the US National Science Foundation under grant IIS-0448542.

The motivation for this book came from the involvement of the first three authors in an earlier grant from the European Union (ESPRIT project METAL [166]). We would like to acknowledge financial support from this grant and the contribution of our colleagues: Hilan Bensusan, Helmut Berrer, Sašo Džeroski, Peter Flach, Johannes Fürnkranz, Melanie Hilario, Jörg Keller, Tom Khabaza, Alexandros Kalousis, Petr Kuba, Rui Leite, Guido Lindner, Reza Nakhaeizadeh, Iain Paterson, Yonghong Peng, Rui Pereira, Johann Petrak, Bernhard Pfahringer, Luboš Popelínský, Ljupčo Todorovski, Dietrich Wettschereck, Gerhard Widmer, Adam Woznica and Bernard Zenko.

We are greatly indebted to several other colleagues for their many comments and suggestions that helped improve earlier versions of this book: Bart Bakker, Theodoros Evgeniou, Tom Heskes, Rich Maclin, Andreas Maurer, Tony Martinez, Massimiliano Pontil, Rajat Raina, André Rossi, Peter Stone, Richard Sutton, Juergen Schmidhuber, Matthew Taylor, Lisa Torrey and Roberto Valerio.

Pavel Brazdil would also like to gratefully acknowledge that Robert Kowalski drew his attention to the topic of metareasoning in the 70's and

to express his gratitude to the late Donald Michie for having accepted the proposal to include metalearning in the preparatory stages of the StatLog project [169] in 1989.

Finally, we are most grateful to our editor at Springer, Ronan Nugent, for his patience, gentle prodding and encouragement throughout this project.

Porto, Portugal; Provo, Houston, USA
July 2008

Pavel Brazdil
Christophe Giraud-Carrier
Carlos Soares
Ricardo Vilalta

Contents

1 Metalearning: Concepts and Systems 1

2 Metalearning for Algorithm Recommendation:
an Introduction 11

3 Development of Metalearning Systems for Algorithm
Recommendation 31

4 Extending Metalearning to Data Mining and KDD 61

5 Combining Base-Learners 73

6 Bias Management in Time-Changing Data Streams 91

7 Transfer of Metaknowledge Across Tasks 109

8 Composition of Complex Systems: Role
of Domain-Specific Metaknowledge 129

References 153

A Terminology 171

B Mathematical Symbols 173

Index 175