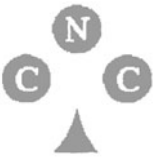
Natural Computing Series



Springer-Verlag Berlin Heidelberg GmbH

Alex A. Freitas

Data Mining and Knowledge Discovery with Evolutionary Algorithms

With 74 Figures and 10 Tables

Springer

Author

Dr. Alex A. Freitas
Computing Laboratory, University of Kent
Canterbury CT2 7NF, UK
A.A.Freitas@ukc.ac.uk

Series Editors

G. Rozenberg (Managing Editor)
Th. Bäck, A.E. Eiben, J.N. Kok, H.P. Spaink

Leiden Center for Natural Computing, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
rozenber@cs.leidenuniv.nl

Library of Congress Cataloging-in-Publication Data
Freitas, Alex. A., 1964–
Data mining and knowledge discovery with evolutionary algorithms/Alex A. Freitas.
p.cm. – (Natural computing series)
Includes bibliographical references and index.
ISBN 978-3-642-07763-0 ISBN 978-3-662-04923-5 (eBook)
DOI 10.1007/978-3-662-04923-5
1. Data mining. 2. Database searching. 3. Computer algorithms. I. Title. II. Series.
QA76.9.D343F722002
006.3–dc21 2002021728

ACM Computing Classification (1998): I., I.2, I.2.6

http://www.springer.de

Originally published by Springer-Verlag Berlin Heidelberg New York in 2002
Softcover reprint of the hardcover 1st edition 2002

Cover design: KünkelLopka, Heidelberg
Typesetting: Steingraeber, Heidelberg
Printed on acid-free paper SPIN: 10986554 45/3111 GF– 54321

This book is dedicated to all the people
who believe that learning is not only
one of the most necessary but also
one of the noblest human activities.

Preface

This book addresses the integration of two areas of computer science, namely data mining and evolutionary algorithms. Both these areas have become increasingly popular in the last few years, and their integration is currently an area of active research.

In essence, data mining consists of extracting valid, comprehensible, and interesting knowledge from data. Data mining is actually an interdisciplinary field, since there are many kinds of methods that can be used to extract knowledge from data. Arguably, data mining mainly uses methods from machine learning (a branch of artificial intelligence) and statistics (including statistical pattern recognition). Our discussion of data mining and evolutionary algorithms is primarily based on machine learning concepts and principles. In particular, in this book we emphasize the importance of discovering comprehensible, interesting knowledge, which the user can potentially use to make intelligent decisions.

In a nutshell, the motivation for applying evolutionary algorithms to data mining is that evolutionary algorithms are robust search methods which perform a global search in the space of candidate solutions (rules or another form of knowledge representation). In contrast, most rule induction methods perform a local, greedy search in the space of candidate rules. Intuitively, the global search of evolutionary algorithms can discover interesting rules and patterns that would be missed by the greedy search.

This book initially presents a comprehensive review of basic concepts from both data mining and evolutionary algorithms, and then it discusses significant advances in the integration of these two areas. It is a self-contained book, explaining both basic concepts and advanced topics in a clear and informal style.

My research on data mining with evolutionary algorithms has been done with the help of many people. I am grateful to my research students for their hard work. They had to carry out a time-consuming iterative process of studying a lot about both data mining and evolutionary algorithms, developing an evolutionary algorithm for data mining, performing several experiments to evaluate the algorithm, analyzing the results, refining the algorithm, etc. That was not an easy task. In particular, I would like to thank (in alphabetical order of first name) Celia C. Bojarczuk, Deborah R. Carvalho, Denise F. Tsunoda, Dieferson L.A. Araujo, Edgar Noda, Fabricio B. Voznika, Fernando E.B. Otero, Gisele L. Pappa, Marcus V. Fidelis, Monique M.S. Silva, Otavio Larsen, Rafael S. Parpinelli, Raul T. Santos, Roberto R.F. Mendes, and Wesley Romao.

I am also grateful to my colleague Dr. Heitor S. Lopes, with whom I have had many interesting discussions about evolutionary algorithms for data mining and have shared the supervision of several of the above-mentioned students.

In addition, I thank an anonymous reviewer for reviewing the manuscript of the book, my colleague Dr. L. Valeria R. Arruda for reviewing the first draft of chapter 11, and Prof. Dr. A.E. Eiben for reviewing chapter 12. I am also grateful to the team at Springer-Verlag for their assistance throughout the project.

Most of the research that led to the writing of this book has been carried out in two universities, both located in Curitiba, Brazil. More precisely, I was with CEFET-PR (Centro Federal de Educacao Tecnologica do Parana) in 1998, and since 1999 I have been with PUC-PR (Pontificia Universidade Catolica do Parana). I thank my colleagues from both institutions for creating a productive, friendly research environment.

Finally, during my research on data mining with evolutionary algorithms I have been partially financially supported by a grant from the Brazilian government's National Council of Scientific and Technological Development (CNPq), process number 300153/98-8.

Curitiba, Brazil
March 2002

Alex A. Freitas

Table of Contents