Studies in Computational Intelligence

Volume 781

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/7092

Jan Kozak

Decision Tree and Ensemble Learning Based on Ant Colony Optimization

Springer

Jan Kozak
Faculty of Informatics and Communication,
Department of Knowledge Engineering
University of Economics in Katowice
Katowice, Poland

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-319-93751-9 ISBN 978-3-319-93752-6 (eBook)
https://doi.org/10.1007/978-3-319-93752-6

Library of Congress Control Number: 2018944347

© Springer International Publishing AG, part of Springer Nature 2019
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, 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.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

My greatest, heartfelt thanks go to my beloved wife Dorota, who stayed with me in good and bad moments—and to my dear daughters: Asia, who showed me, that there is no such thing as impossible and that I should never give up, and Ola, who taught me how much we can achieve in life by helping others.

Thank you…

Preface

This book, as its title suggests, is devoted to decision trees and ensemble learning based on ant colony optimization. Accordingly, it not only concerns the aforementioned important topics in the area of machine learning and combinatorial optimization, but combines them into one. It is this combination that was decisive for choosing the material to be included in the book and determining its order of presentation.

Decision trees are a popular method of classification as well as of knowledge representation. As conceptually straightforward, they are very appealing and have proved very useful due to their computational efficiency and generally acceptable effectiveness. At the same time, they are easy to implement as the building blocks of an ensemble of classifiers—in particular, decision forests. It is worth mentioning that decision forests are now considered to be the best available off-the-shelf classifiers. Admittedly, however, the task of constructing a decision tree is a very complex process if one aims at the optimality, or near-optimality, of the tree. Indeed, we usually agree to proceed in a heuristic way, making multiple decisions: one at each stage of the multistage tree construction procedure, but each of them obtained independently of any other. However, construction of a minimal, optimal decision tree is an NP-complete problem. The underlying reason is that local decisions are in fact interdependent and cannot be found in the way suggested earlier.

The good results typically achieved by the ant colony optimization algorithms when dealing with combinatorial optimization problems suggest the possibility of using that approach effectively also for constructing decision trees. The underlying rationale is that both problem classes can be presented as graphs. This fact leads to the possibility of considering a larger spectrum of solutions than those based on the heuristic mentioned earlier. Namely, when searching for a solution, ant colony optimization algorithms allow for using additional knowledge acquired from the pheromone trail values produced by such algorithms. Moreover, ant colony optimization algorithms can be used to advantage when building ensembles of classifiers.

This book is a combination of a research monograph and a textbook. It can be used in graduate courses, but should also be of interest to researchers, both specialists in machine learning and those applying machine learning methods to cope with problems from any field of R&D. The topics included in the book are discussed thoroughly. Each of them is introduced in such a way as to make it accessible to a computer science or engineering graduate student. All methods included are discussed from many angles, in particular regarding their applicability and the ways to choose their parameters. In addition, many unpublished results are presented.

The book is divided into two parts, preceded by an introduction to machine learning and swarm intelligence. The first part discusses decision tree learning based on ant colony optimization. It includes an introduction to evolutionary computing techniques for data mining (Chap. 2). Chapter 3 describes the most popular ant colony algorithms used for learning decision trees. Chapter 4 introduces some modifications of the ant colony decision tree algorithm. Examples of practical applications of the methods discussed earlier are presented in Chap. 5.

The second part of the book discusses ensemble learning based on ant colony optimization. It begins (Chap. 6) with an introduction presenting known solutions of a similar type—more precisely, evolutionary computing techniques in ensemble learning—as well as a literature review. Chapter 7 includes formal definitions and example applications of ensembles based on random forests. Chapter 8 introduces an adaptive approach to building a decision forest with ant colony optimization. The book concludes with some final remarks and suggestions for future research.

Let me end with mentioning the persons whose advice, help and support were indispensable for writing this book. Special thanks are due to Prof. Jacek Koronacki for the possibility of collaboration, his extraordinary commitment, fruitful discussions and all comments. Similar thanks to Prof. Beata Konikowska, who showed unusual meticulousness and gave me many valuable remarks and pieces of advice. I would also like to thank Prof. Mikhail Moshkov for his continuing comments and long discussions. Sincere thanks are due to Dr. Przemysaw Juszczuk, who has been deeply involved in creating this book.

Katowice, Poland Jan Kozak
October 2017

Contents