Evolutionary Learning: Advances in Theories and Algorithms

Zhi-Hua Zhou · Yang Yu ·
Chao Qian

Evolutionary Learning: Advances in Theories and Algorithms

Springer

Zhi-Hua Zhou
Nanjing University
Nanjing, Jiangsu, China

Yang Yu
Nanjing University
Nanjing, Jiangsu, China

Chao Qian
Nanjing University
Nanjing, Jiangsu, China

ISBN 978-981-13-5955-2 ISBN 978-981-13-5956-9 (eBook)
https://doi.org/10.1007/978-981-13-5956-9

© Springer Nature Singapore Pte Ltd. 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, expressed 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.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Around the year of 2001, the first author of this monograph, Zhou, developed with his collaborators a powerful selective ensemble approach. This approach is able to produce small-sized ensembles with generalization performance even stronger than that of full-sized ones by exploiting genetic algorithm, a popularly used evolutionary algorithm (EA). Zhou realized that EAs are powerful optimization techniques that can be well useful in various machine learning tasks. At that time, however, EAs were almost purely heuristic, not favored by the machine learning community with strong theoretical flavor. Impressed by the successes of EAs in applications, Zhou believed there must be some theoretical explanations behind their mysteries and decided to start this track of research. In 2004, the second author of this monograph, Yu, finished his bachelor thesis on selective ensemble under the supervision of Zhou. Yu then joined Zhou as a PhD student taking theoretical aspects of EAs as his thesis topic, and obtained PhD degree in 2011. In 2009, Zhou accepted the third author of this monograph, Qian, as his PhD student with theories and algorithms of evolutionary learning for his thesis topic, and Qian obtained his PhD degree in 2015. Overall, the majority of contents in this monograph are research results achieved by the authors during the past two decades.

This monograph is composed of four parts. Part I briefly introduces evolutionary learning and some preliminaries. Aiming at analyzing the running time complexity and the approximation ability, the two essential theoretical aspects of EAs, Part II presents two general approaches for deriving running time bounds and a general framework for characterizing the approximation performance. These results serve as general tools for attaining many theoretical results reported in the remainder of the monograph. Part III presents a series of theoretical studies, mainly about the influence of major factors, such as recombination, solution representation, inaccurate fitness evaluation, population, etc., on the performance of EAs. In Part IV, the authors come back to selective ensemble, their original motivating task for studying EAs, and present algorithms behaving at least as well as state-of-the-art

selective ensemble algorithms. The authors then study subset selection, a more general problem widely occurring in various machine learning tasks, and present a series of evolutionary learning algorithms with bounded approximation guarantees.

The authors hope that the general theoretical tools presented in Part II can be found useful for interested readers to explore theoretical foundations of evolutionary learning, the theoretical results presented in Part III can help readers to get more understanding about behaviors of evolutionary learning processes and offer some insight for algorithm design, and the algorithms presented in Part IV can be found useful in real-world applications of machine learning.

The authors want to thank their families, friends and collaborators. Gratitude goes to Celine Chang and Alfred Hofmann who encouraged authors to publish this monograph with Springer.

The authors
February 2019

Notations

$\mathbb{R}$	real number
$\mathbb{N}$	integer
$(\cdot)^{+}$	positive, $(\cdot)$ can be $\mathbb{R}$ or $\mathbb{N}$
$(\cdot)^{0+}$	non-negative, $(\cdot)$ can be $\mathbb{R}$ or $\mathbb{N}$
x	variable
$\boldsymbol{x}$	vector
$(\cdot, \cdot, \ldots, \cdot)$	row vector
$\mathbf{0}, \mathbf{1}$	all-0s and all-1s vectors
$\lvert \cdot \rvert_0, \lvert \cdot \rvert_1$	number of 0s and 1s of a vector
$\{0, 1\}^n$	Boolean vector space
$\mathbf{X}$	matrix
$(\cdot)^{\mathrm{T}}$	transpose of a vector/matrix
X	set
$\{\cdot, \cdot, \cdots, \cdot\}$	set by enumeration
$[n]$	set $\{1, 2, \ldots, n\}$
$\lvert \cdot \rvert$	cardinality of a set
2^X	power set of X, which consists of all subsets of X
$X \setminus Y$	complement of Y in X, which consists of elements in X but not in Y

x	state of a Markov chain
$\mathcal{X}$	state space of a Markov chain
$\mathrm{P}(\cdot), \mathrm{P}(\cdot \mid \cdot)$	probability and conditional probability
π	probability distribution
f	function
$\mathbb{E}_{\cdot\sim\pi}[f(\cdot)], \mathbb{E}_{\cdot\sim\pi}[f(\cdot) \mid \cdot]$	expectation and conditional expectation of $f(\cdot)$ under distribution π, simplified as $\mathbb{E}[f(\cdot)], \mathbb{E}[f(\cdot) \mid \cdot]$ when the meaning is clear
$\mathbb{I}(\cdot)$	indicator function which takes 1 if $\cdot$ is true, and 0 otherwise
$\lfloor\cdot\rfloor, \lceil\cdot\rceil$	floor and ceiling functions which take the greatest/least integer less/greater than or equal to a real number
OPT	optimal function value
H_n	n-th harmonic number, i.e., $\sum_{i=1}^{n}(1/i)$
$\forall$	for all
$\exists, \nexists$	there exists, does not exist

Contents

Part I INTRODUCTION

1 Introduction 3
1.1 Machine Learning 3
1.2 Evolutionary Learning 4
1.3 Multi-objective Optimization 6
1.4 Organization of the Book 9

2 Preliminaries 11
2.1 Evolutionary Algorithms 11
2.2 Pseudo-Boolean Functions 15
2.3 Running Time Complexity 17
2.4 Markov Chain Modeling 19
2.5 Analysis Tools 21

Part II ANALYSIS METHODOLOGY

3 Running Time Analysis: Convergence-based Analysis 29
3.1 Convergence-based Analysis: Framework 30
3.2 Convergence-based Analysis: Application Illustration 34
3.3 Summary 39

4 Running Time Analysis: Switch Analysis 41
4.1 Switch Analysis: Framework 41
4.2 Switch Analysis: Application Illustration 46
4.3 Summary 50

5 Running Time Analysis: Comparison and Unification 51
5.1 Analysis Approaches: Formalization 51
5.2 Switch Analysis vs. Fitness Level 53

5.3 Switch Analysis vs. Drift Analysis 57
5.4 Switch Analysis vs. Convergence-based Analysis 61
5.5 Analysis Approaches: Unification 65
5.6 Summary 67

6 Approximation Analysis: SEIP 69
6.1 SEIP: Framework 70
6.2 SEIP: Application Illustration 77
6.3 Summary 80

Part III THEORETICAL PERSPECTIVES

7 Boundary Problems of EAs 83
7.1 Boundary Problem Identification 84
7.2 Case Study 87
7.3 Summary 92

8 Recombination 93
8.1 Recombination and Mutation 95
8.2 Recombination Enabled MOEAs 97
8.3 Case Study 100
8.4 Empirical Verification 106
8.5 Summary 108

9 Representation 109
9.1 Genetic Programming Representation 111
9.2 Case Study: Maximum Matchings 113
9.3 Case Study: Minimum Spanning Trees 119
9.4 Empirical Verification 125
9.5 Summary 128

10 Inaccurate Fitness Evaluation 129
10.1 Noisy Optimization 130
10.2 Influence of Noisy Fitness 132
10.3 Denoise by Threshold Selection 136
10.4 Denoise by Sampling 141
10.5 Empirical Verification 148
10.6 Summary 152

11 Population 155
11.1 Influence of Population 156
11.2 Robustness of Population against Noise 160
11.3 Summary 172

12 Constrained Optimization 175
12.1 Usefulness of Infeasible Solutions 177
12.2 Effectiveness of Pareto Optimization 183
12.3 Summary 194

Part IV LEARNING ALGORITHMS

13 Selective Ensemble 197
13.1 Selective Ensemble 198
13.2 The POSE Algorithm 200
13.3 Theoretical Analysis 202
13.4 Empirical Study 208
13.5 Summary 214

14 Subset Selection 215
14.1 Subset Selection 216
14.2 The POSS Algorithm 221
14.3 Theoretical Analysis 222
14.4 Empirical Study 226
14.5 Summary 230

15 Subset Selection: k-Submodular Maximization 233
15.1 Monotone k-Submodular Function Maximization 235
15.2 The POkSS Algorithm 238
15.3 Theoretical Analysis 241
15.4 Empirical Study 246
15.5 Summary 254

16 Subset Selection: Ratio Minimization 255
16.1 Ratio Minimization of Monotone Submodular Functions . . . 256
16.2 The PORM Algorithm 259
16.3 Theoretical Analysis 261
16.4 Empirical Study 266
16.5 Summary 268

17 Subset Selection: Noise 269
17.1 Noisy Subset Selection 270
17.2 The PONSS Algorithm 275
17.3 Theoretical Analysis 277
17.4 Empirical Study 280
17.5 Summary 283

18 Subset Selection: Acceleration 285
18.1 The PPOSS Algorithm 285
18.2 Theoretical Analysis 287
18.3 Empirical Study 291
18.4 Summary 293

Appendix 295
A.1 Proofs in Chapter 4 295
A.2 Proofs in Chapter 5 298
A.3 Proofs in Chapter 8 306
A.4 Proofs in Chapter 9 310
A.5 Proofs in Chapter 10 312
A.6 Proofs in Chapter 11 334
A.7 Proofs in Chapter 12 335
A.8 Proofs in Chapter 15 339
A.9 Proofs in Chapter 17 340

References 343