

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Elena Marchiori Jason H. Moore
Jagath C. Rajapakse (Eds.)

Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics

5th European Conference, EvoBIO 2007
Valencia, Spain, April 11-13, 2007
Proceedings

Volume Editors

Elena Marchiori
VU University of Amsterdam, IBIVU
Department of Computer Science
de Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
E-mail: elena@cs.vu.nl

Jason H. Moore
Dartmouth-Hitchcock Medical Center
Computational Genetics Laboratory
706 Rubin Building, HB 7937, One Medical Center Dr., Lebanon, NH 03756, USA
E-mail: jason.h.moore@dartmouth.edu

Jagath C. Rajapakse
Nanyang Technological University
School of Computer Engineering
Blk N4-2a05, 50 Nanyang Avenue, Singapore 639798
E-mail: asjagath@ntu.edu.sg

Cover illustration: Morphogenesis series #12 by Jon McCormack, 2006

Library of Congress Control Number: 2007923724

CR Subject Classification (1998): D.1, F.1-2, J.3, I.5, I.2

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN	0302-9743
ISBN-10	3-540-71782-X Springer Berlin Heidelberg New York
ISBN-13	978-3-540-71782-9 Springer Berlin Heidelberg New York

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, re-use of illustrations, recitation, broadcasting, reproduction on microfilms 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.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12044597 06/3180 5 4 3 2 1 0

Preface

The field of bioinformatics has two main objectives: the creation and maintenance of biological databases, and the discovery of knowledge from life sciences data in order to unravel the mysteries of biological function, leading to new drugs and therapies for human disease. Life sciences data come in the form of biological sequences, structures, pathways, or literature. One major aspect of discovering biological knowledge is to search, predict, or model specific patterns present in a given dataset and then to interpret those patterns. Computer science methods such as evolutionary computation, machine learning, and data mining all have a great deal to offer the field of bioinformatics. The goal of the Fifth European Conference on Evolutionary Computation, Machine Learning, and Data Mining in Bioinformatics (EvoBIO 2007) was to bring experts in computer science together with experts in bioinformatics and the biological sciences to explore new and novel methods for solving complex biological problems.

The fifth EvoBIO conference was held in Valencia, Spain during April 11-13, 2007 at the Universidad Politecnica de Valencia. EvoBIO 2007 was held jointly with the Tenth European Conference on Genetic Programming (EuroGP 2007), the Seventh European Conference on Evolutionary Computation in Combinatorial Optimisation (EvoCOP 2007), and the Evo Workshops. Collectively, the conferences and workshops are organized under the name Evo* (www.evostar.org).

EvoBIO, held annually as a workshop since 2003, became a conference in 2007 and it is now the premiere European event for those interested in the interface between evolutionary computation, machine learning, data mining, bioinformatics, and computational biology. All papers in this book were presented at EvoBIO 2007 and responded to a call for papers that included topics of interest such as biomarker discovery, cell simulation and modeling, ecological modeling, fluxomics, gene networks, biotechnology, metabolomics, microarray analysis, phylogenetics, protein interactions, proteomics, sequence analysis and alignment, and systems biology. A total of 60 papers were submitted to the conference for double-blind peer-review. Of those, 28 (46.7%) were accepted.

We would first and foremost like to thank all authors who spent time and effort to make important contributions to this book. We would like to thank the members of the Program Committee for their expert evaluation of the submitted papers. Moreover, we would like to thank Jennifer Willies for her tremendous administrative help and coordination, Anna Isabel Esparcia-Alcázar for serving as the Local Chair, Leonardo Vanneschi for serving as Evo* Publicity Chair, Marc Schoenauer and the MyReview team (<http://myreview.lri.fr/>) for the conference management system.

We would also like to acknowledge the following organizations. The Universidad Politécnica de Valencia, Spain for their institutional and financial support, and for providing premises and administrative assistance; the Instituto

Tecnológico de Informática in Valencia, for cooperation and help with local arrangements; the Spanish Ministerio de Educación y Ciencia, for their financial support; and the Centre for Emergent Computing at Napier University in Edinburgh, Scotland for administrative support and event coordination.

Finally, we hope that you will consider contributing to EvoBIO 2008.

February 2007

Elena Marchiori
Jason H. Moore
Jagath C. Rajapakse

Organization

EvoBIO 2007 was organized by Evo* (www.evostar.org).

Program Chairs

Elena Marchiori (IBIVU, VU University Amsterdam, The Netherlands)
Jason H. Moore (Dartmouth Medical School in Lebanon, NH, USA)
Jagath C. Rajapakse (Nanyang Technological University, Singapore)

General Chairs

David W. Corne (Heriot-Watt University, Edinburgh, UK)
Elena Marchiori (IBIVU, VU University Amsterdam, The Netherlands)

Steering Committee

David W. Corne (Heriot-Watt University, Edinburgh, UK)
Elena Marchiori (IBIVU, VU University Amsterdam, The Netherlands)
Carlos Cotta (University of Malaga, Spain)
Jason H. Moore (Dartmouth Medical School in Lebanon, NH, USA)
Jagath C. Rajapakse (Nanyang Technological University, Singapore)

Program Committee

Jesus S. Aguilar-Ruiz (Spain)	Elena Marchiori (The Netherlands)
Francisco J. Azuaje (UK)	Andrew Martin (UK)
Wolfgang Banzhaf (Canada)	Jason Moore (USA)
Jacek Blazewicz (Poland)	Pablo Moscato (Australia)
Marius Codrea (The Netherlands)	Jagath Rajapakse (Singapore)
Dave Corne (UK)	Menaka Rajapakse (Singapore)
Carlos Cotta (Spain)	Michael Raymer (USA)
Alex Freitas (UK)	Vic J. Rayward-Smith (UK)
Gary Fogel (USA)	Jem Rowland (UK)
James Foster (USA)	Marylyn Ritchie (USA)
Rosalba Giugno (Italy)	Ugur Sezerman (Turkey)
Raul Giraldez (Spain)	El-Ghazali Talbi (France)
Jin-Kao Hao (France)	Andrea Tettamanzi (Italy)
Antoine van Kampen (The Netherlands)	Janet Wiles (Australia)
Natalio Krasnogor (UK)	Andreas Zell (Germany)
Ying Liu (USA)	Eckart Zitzler (Switzerland)

Table of Contents

Identifying Regulatory Sites Using Neighborhood Species	1
<i>Claudia Angelini, Luisa Cutillo, Italia De Feis, Richard van der Wath, and Pietro Lio'</i>	
Genetic Programming and Other Machine Learning Approaches to Predict Median Oral Lethal Dose (LD ₅₀) and Plasma Protein Binding Levels (%PPB) of Drugs	11
<i>Francesco Archetti, Stefano Lanzeni, Enza Messina, and Leonardo Vanneschi</i>	
Hypothesis Testing with Classifier Systems for Rule-Based Risk Prediction	24
<i>Flavio Baronti and Antonina Starita</i>	
Robust Peak Detection and Alignment of nanoLC-FT Mass Spectrometry Data	35
<i>Marius C. Codrea, Connie R. Jiménez, Sander Piersma, Jaap Heringa, and Elena Marchiori</i>	
One-Versus-One and One-Versus-All Multiclass SVM-RFE for Gene Selection in Cancer Classification	47
<i>Kai-Bo Duan, Jagath C. Rajapakse, and Minh N. Nguyen</i>	
Understanding Signal Sequences with Machine Learning	57
<i>Jean-Luc Falcone, Renée Kreuter, Dominique Belin, and Bastien Chopard</i>	
Targeting Differentially Co-regulated Genes by Multiobjective and Multimodal Optimization	68
<i>Oscar Harari, Cristina Rubio-Escudero, and Igor Zwir</i>	
Modeling Genetic Networks: Comparison of Static and Dynamic Models	78
<i>Cristina Rubio-Escudero, Oscar Harari, Oscar Cordón, and Igor Zwir</i>	
A Genetic Embedded Approach for Gene Selection and Classification of Microarray Data	90
<i>Jose Crispin Hernandez Hernandez, Béatrice Duval, and Jin-Kao Hao</i>	
Modeling the Shoot Apical Meristem in <i>A. thaliana</i> : Parameter Estimation for Spatial Pattern Formation	102
<i>Tim Hohm and Eckart Zitzler</i>	

Evolutionary Search for Improved Path Diagrams	114
<i>Kim Laurio, Thomas Svensson, Mats Jirstrand, Patric Nilsson, Jonas Gamalielsson, and Björn Olsson</i>	
Simplifying Amino Acid Alphabets Using a Genetic Algorithm and Sequence Alignment	122
<i>Jacek Lenckowski and Krzysztof Walczak</i>	
Towards Evolutionary Network Reconstruction Tools for Systems Biology	132
<i>Thorsten Lenser, Thomas Hinze, Bashar Ibrahim, and Peter Dittrich</i>	
A Gaussian Evolutionary Method for Predicting Protein-Protein Interaction Sites	143
<i>Kang-Ping Liu and Jinn-Moon Yang</i>	
Bio-mimetic Evolutionary Reverse Engineering of Genetic Regulatory Networks	155
<i>Daniel Marbach, Claudio Mattiussi, and Dario Floreano</i>	
Tuning ReliefF for Genome-Wide Genetic Analysis	166
<i>Jason H. Moore and Bill C. White</i>	
Dinucleotide Step Parameterization of Pre-miRNAs Using Multi-objective Evolutionary Algorithms	176
<i>Jin-Wu Nam, In-Hee Lee, Kyu-Baek Hwang, Seong-Bae Park, and Byoung-Tak Zhang</i>	
Amino Acid Features for Prediction of Protein-Protein Interface Residues with Support Vector Machines	187
<i>Minh N. Nguyen, Jagath C. Rajapakse, and Kai-Bo Duan</i>	
Predicting HIV Protease-Cleavable Peptides by Discrete Support Vector Machines	197
<i>Carlotta Orsenigo and Carlo Vercellis</i>	
Inverse Protein Folding on 2D Off-Lattice Model: Initial Results and Perspectives	207
<i>David Pelta and Alberto Carrascal</i>	
Virtual Error: A New Measure for Evolutionary Biclustering	217
<i>Beatriz Pontes, Federico Divina, Raúl Giraldez, and Jesús S. Aguilar-Ruiz</i>	
Characterising DNA/RNA Signals with Crisp Hypermotifs: A Case Study on Core Promoters	227
<i>Carey Pridgeon and David Corne</i>	

Evaluating Evolutionary Algorithms and Differential Evolution for the Online Optimization of Fermentation Processes	236
<i>Miguel Rocha, José P. Pinto, Isabel Rocha, and Eugénio C. Ferreira</i>	
The Role of a Priori Information in the Minimization of Contact Potentials by Means of Estimation of Distribution Algorithms	247
<i>Roberto Santana, Pedro Larrañaga, and Jose A. Lozano</i>	
Classification of Cell Fates with Support Vector Machine Learning	258
<i>Ofer M. Shir, Vered Raz, Roeland W. Dirks, and Thomas Bäck</i>	
Reconstructing Linear Gene Regulatory Networks	270
<i>Jochen Supper, Christian Spieth, and Andreas Zell</i>	
Individual-Based Modeling of Bacterial Foraging with Quorum Sensing in a Time-Varying Environment	280
<i>W.J. Tang, Q.H. Wu, and J.R. Saunders</i>	
Substitution Matrix Optimisation for Peptide Classification	291
<i>David C. Trudgian and Zheng Rong Yang</i>	
Author Index	301