Lecture Notes in Computer Science 6623

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

Alfred Kobsa
University of California, Irvine, CA, 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
TU Dortmund University, Germany

Madhu Sudan
Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Clara Pizzuti Marylyn D. Ritchie
Mario Giacobini (Eds.)

Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics

9th European Conference, EvoBIO 2011
Torino, Italy, April 27-29, 2011
Proceedings

Springer

Volume Editors

Clara Pizzuti
Institute for High-Performance Computing and Networking (ICAR)
Italian National Research Council (CNR)
Via P. Bucci 41C, 87036 Rende (CS), Italy
E-mail: pizzuti@icar.cnr.it

Marylyn D. Ritchie
Vanderbilt University
Center for Human Genetics Research
519 Light Hall, Nashville, TN 37232, USA
E-mail: ritchie@chgr.mc.vanderbilt.edu

Mario Giacobini
University of Torino
Department of Animal Production Epidemiology and Ecology
Via Leonardo da Vinci 44, 10095 Grugliasco (TO), Italy
E-mail: mario.giacobini@unito.it

Cover illustration:
"Globosphere" by Miguel Nicolau and Dan Costelloe (2010),
University of Dublin, Ireland

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-20388-6 e-ISBN 978-3-642-20389-3
DOI 10.1007/978-3-642-20389-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011924465

CR Subject Classification (1998): J.3, H.2.8, E.1, I.2, F.1, F.2.1

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

© Springer-Verlag Berlin Heidelberg 2011
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.
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.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

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 information in a given dataset in order to generate new interesting knowledge. 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 9th European Conference on Evolutionary Computation, Machine Learning, and Data Mining in Bioinformatics (EvoBIO 2011) was to bring together experts in these fields in order to discuss new and novel methods for tackling complex biological problems.

The 9th EvoBIO conference was held in Torino, Italy, during April 27–29, 2011 at the University of Torino. EvoBIO 2011 was held jointly with the 14th European Conference on Genetic Programming (EuroGP 2011), the 11th European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP 2011), and the conference on the applications of evolutionary computation, EvoAPPLICATIONS. Collectively, the conferences 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 premier 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 2011 in oral or poster presentations 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. After peer review, we accepted 12 papers for oral presentation and 7 for poster presentation.

We would first and foremost like to thank all authors who spent time and effort to produce interesting contributions to this book. We would like to acknowledge Mario Giacobini, of the University of Torino, for his outstanding work editing the EvoBIO 2011 volume. We would like to thank the members of the Program Committee for their expert evaluation of the submitted papers, Jennifer Willies from Edinburgh Napier University, for her tremendous administrative help and coordination, and other personnel at the Centre for Emergent Computing at Edinburgh Napier University, UK, for assisting with coordination. We would like to thank Penousal Machado, University of Coimbra, for his fantastic work as the Publicity Chair. We would like to thank Marc Schoenauer from

INRIA in France, for his continued support and assistance with the MyReview conference management system. We would also like to extend special thanks to Mario Giacobini, University of Torino, for his tremendous work as the local organizer. Moreover, we would like to thank the following institutes: the Human Genetics Foundation of Torino (HuGeF), the School for Biotechnologies and the Molecular Biotechnology Center of the University of Torino, the Museum of Human Anatomy "Luigi Rolando," and also the Museum of Criminal Anthropology "Cesare Lombroso" for local support.

We want to especially acknowledge the invited speakers that during the conference days gave two very interesting and inspirational talks: Jean-Pierre Changeux, professor emeritus at the Collége de France and at the Institut Pasteur, France, and Craig Reynolds, senior researcher at Sony Computer Entertainment, USA.

Finally, we hope that you will consider contributing to EvoBIO 2012.

April 2011

Clara Pizzuti
Marylyn D. Ritchie

Organization

EvoBIO 2011, together with EuroGP 2011, EvoCOP 2011, and EvoAPPLICATIONS 2011, was part of EVO* 2011, Europe's premier co-located events in the field of evolutionary computing.

Program Chairs

Clara Pizzuti	Institute for High Performance Computing and Networking National Research Council (ICAR-CNR), Italy
Marylyn D. Ritchie	Vanderbilt University in Nashville, USA

Local Chair

Mario Giacobini	University of Torino, Italy

Publicity Chair

Penousal Machado	University of Coimbra, Portugal

Proceedings Chair

Mario Giacobini	University of Torino, Italy

Steering Committee

David W. Corne	Heriot-Watt University, Edinburgh, UK
Elena Marchiori	Radboud University, Nijmegen, 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	Pablo de Olavide University, Spain
Wolfgang Banzhaf	Memorial University of Newfoundland, Canada
Jacek Blazewicz	Poznan University of Technology, Poland
Erik Boczko,	Vanderbilt University, USA
William Bush	Vanderbilt University, USA
Carlos Cotta	University of Malaga, Spain

Federico Divina	Pablo de Olavide University, Spain
Jitesh Dundas	Edencore Technologies,USA
Gary Fogel	Natural Selection, Inc., USA
Alex Freitas	University of Kent, UK
Mario Giacobini	Polotecnico di Torino, Italy
Raffaele Giancarlo	Università di Palermo, Italy
Rosalba Giugno	Università di Catania, Italy
Jin-Kao Hao	University of Angers, France
Tom Heskes	Radboud University Nijmegen, The Netherlands
Zhenyu Jia	University of California, Irvine, USA
Mehmet Koyuturk	Case Western Reserve University, USA
Michael Lones	University of York, UK
Bob MacCallum	Imperial College London, UK
Elena Marchiori	Radboud University, Nijmegen, The Netherlands
Andrew Martin	University College London, UK
Brett McKinney	University of Alabama, Birmingham, USA
Pablo Moscato	The University of Newcastle, Australia
Alison Motsinger-Reif	North Carolina State University, USA
Vincent Moulton	University of East Anglia, UK
See-Kiong Ng	Institute for Infocomm Research, Singapore
Carlotta Orsenigo	Politecnico di Milano, Italy
Clara Pizzuti	ICAR-CNR, Italy
Michael Raymer	Wright State University, USA
Marylyn Ritchie	Vanderbilt University, USA
Raul Giraldez Rojo	Pablo de Olavide University, Spain
Simona Rombo	Università della Calabria Italy
Marc Schoenauer	LRI- Université Paris-Sud, France
Ugur Sezerman	Sabanci University, Turkey
Marc L. Smith	Vassar College, USA
El-Ghazali Talbi	Université des Sciences et Technologies de Lille, France
Alfonso Urso	ICAR-CNR, Italy
Antoine van Kampen	Universiteit van Amsterdam, The Netherlands
Tiffani Williams	Texas A&M University, USA
Andreas Zell	University of Tübingen, Germany
Zhongming Zhao	Vanderbilt University, USA
Blaz Zupan	University of Ljubljana, Slovenia

Sponsoring Institutions

- Human Genetics Foundation of Torino (HuGeF)
- School for Biotechnologies of the University of Torino
- Molecular Biotechnology Center of the University of Torino
- Museum of Human Anatomy “Luigi Rolando” of Torino
- Museum of Criminal Anthropology “Cesare Lombroso” of Torino
- The Centre for Emergent Computing, Edinburgh Napier University, UK

Table of Contents

Effect of Using Varying Negative Examples in Transcription Factor Binding Site Predictions 1
Faisal Rezwan, Yi Sun, Neil Davey, Rod Adams, Alistair G. Rust, and Mark Robinson

A New Evolutionary Gene Regulatory Network Reverse Engineering Tool 13
Antonella Farinaccio, Leonardo Vanneschi, Paolo Provero, Giancarlo Mauri, and Mario Giacobini

ML-Consensus: A General Consensus Model for Variable-Length Transcription Factor Binding Sites 25
Saad Quader, Nathan Snyder, Kevin Su, Ericka Mochan, and Chun-Hsi Huang

Applying Linear Models to Learn Regulation Programs in a Transcription Regulatory Module Network 37
Jianlong Qi, Tom Michoel, and Gregory Butler

ATHENA Optimization: The Effect of Initial Parameter Settings across Different Genetic Models 48
Emily R. Holzinger, Scott M. Dudek, Eric C. Torstenson, and Marylyn D. Ritchie

Validating a Threshold-Based Boolean Model of Regulatory Networks on a Biological Organism 59
Christian Darabos, Ferdinando Di Cunto, Marco Tomassini, Jason H. Moore, Paolo Provero, and Mario Giacobini

A Nearest Neighbour-Based Approach for Viral Protein Structure Prediction 69
Gualberto Asencio Cortés, Jesús S. Aguilar-Ruiz, and Alfonso E. Márquez Chamorro

Annotated Stochastic Context Free Grammars for Analysis and Synthesis of Proteins 77
Eva Sciacca, Salvatore Spinella, Dino Ienco, and Paola Giannini

Finding Motifs in DNA Sequences Applying a Multiobjective Artificial Bee Colony (MOABC) Algorithm 89
David L. González-Álvarez, Miguel A. Vega-Rodríguez, Juan A. Gómez-Pulido, and Juan M. Sánchez-Pérez

An Evolutionary Approach for Protein Contact Map Prediction 101
Alfonso E. Márquez Chamorro, Federico Divina, Jesús S. Aguilar-Ruiz, and Gualberto Asencio Cortés

Multi-Neighborhood Search for Discrimination of Signal Peptides and Transmembrane Segments 111
Sami Laroum, Béatrice Duval, Dominique Tessier, and Jin-Kao Hao

Approximation of Graph Kernel Similarities for Chemical Graphs by Kernel Principal Component Analysis 123
Georg Hinselmann, Andreas Jahn, Nikolas Fechner, Lars Rosenbaum, and Andreas Zell

Posters

Experimental Approach for Bacterial Strains Characterization 135
Fabien Chhel, Adrien Goëffon, Frédéric Lardeux, Frédéric Saubion, Gilles Hunault, and Tristan Boureau

Do Diseases Spreading on Bipartite Networks Have Some Evolutionary Advantage? 141
Luca Ferreri, Ezio Venturino, and Mario Giacobini

Genetic Algorithm Optimization of Force Field Parameters: Application to a Coarse-Grained Model of RNA 147
Filip Leonarski, Fabio Trovato, Valentina Tozzini, and Joanna Trylska

A Decision Tree-Based Method for Protein Contact Map Prediction 153
Cosme Ernesto Santiesteban Toca, Alfonso E. Márquez Chamorro, Gualberto Asencio Cortés, and Jesus S. Aguilar-Ruiz

A Comparison of Machine Learning Methods for the Prediction of Breast Cancer 159
Sara Silva, Orlando Anunciação, and Marco Lotz

An Automatic Identification and Resolution System for Protein-Related Abbreviations in Scientific Papers 171
Paolo Atzeni, Fabio Polticelli, and Daniele Toti

Protein Complex Discovery from Protein Interaction Network with High False-Positive Rate 177
Yunku Yeu, Jaegyoon Ahn, Youngmi Yoon, and Sanghyun Park

Author Index 183