

Lecture Notes in Bioinformatics

5780

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand
T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff
R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Visakan Kadiramanathan
Guido Sanguinetti Mark Girolami
Mahesan Niranjan Josselin Noirel (Eds.)

Pattern Recognition in Bioinformatics

4th IAPR International Conference, PRIB 2009
Sheffield, UK, September 7-9, 2009
Proceedings

Series Editors

Sorin Istrail, Brown University, Providence, RI, USA

Pavel Pevzner, University of California, San Diego, CA, USA

Michael Waterman, University of Southern California, Los Angeles, CA, USA

Volume Editors

Visakan Kadiramanathan

Guido Sanguinetti

Josselin Noirel

University of Sheffield

Mappin Street, Sheffield, S1 3JD, UK

E-mail: {visakan,g.sanguinetti,j.noirel}@sheffield.ac.uk

Mark Girolami

University of Glasgow

Glasgow, G12 8QQ, UK

E-mail: girolami@dcs.gla.ac.uk

Mahesan Niranjan

University of Southampton

Southampton SO17 1BJ, UK

E-mail: mn@ecs.soton.ac.uk

Library of Congress Control Number: 2009933351

CR Subject Classification (1998): J.3, I.5, F.2.2, I.2, H.3.3, H.2.8

LNCS Sublibrary: SL 8 – Bioinformatics

ISSN 0302-9743

ISBN-10 3-642-04030-6 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-04030-6 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.com

© Springer-Verlag Berlin Heidelberg 2009

Printed in Germany

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

Printed on acid-free paper SPIN: 12747078 06/3180 5 4 3 2 1 0

Preface

The Pattern Recognition in Bioinformatics (PRIB) meeting was established in 2006 under the auspices of the International Association for Pattern Recognition (IAPR) to create a focus for the development and application of pattern recognition techniques in the biological domain. PRIB's aim to explore the full spectrum of pattern recognition application was reflected in the breadth of techniques represented in this year's submissions and in this book. These range from image analysis for biomedical data to systems biology.

We were fortunate to have invited speakers of the highest calibre delivering keynotes at the conference. These were Pierre Baldi (UC Irvine), Alvis Brazma (EMBL-EBI), Gunnar Rätsch (MPI Tübingen) and Michael Unser (EPFL). We acknowledge support of the EU FP7 Network of Excellence PASCAL2 for partially funding the invited speakers.

Immediately prior to the conference, we hosted half day of tutorial lectures, while a special session on "Machine Learning for Integrative Genomics" was held immediately after the main conference. During the conference, a poster session was held with further discussion.

We would like once again to thank all the authors for the high quality of submissions, as well as Yorkshire South and the University of Sheffield for providing logistical help in organizing the conference. Finally, we would like to thank Springer for their help in assembling this proceedings volume and for the continued support of PRIB.

July 2009

Mark Girolami
Visakan Kadirkamanathan
Mahesan Niranjan
Josselin Noirel
Guido Sanguinetti

Organization

International Program Committee

Shandar Ahmed	National Institute of Biomedical Innovation, Japan
Jesús S. Aguilar-Ruiz	Escuela Politécnica Superior, Seville, Spain
Tatsuya Akutsu	Kyoto University, Japan
Sanghamitra Bandyopadhyay	Indian Statistical Institute, India
Sebastian Böcker	Friedrich-Schiller-Universität, Jena, Germany
Rainer Breitling	University of Groningen, The Netherlands
Nicolas Brunel	CNRS, Paris, France
Colin Campbell	University of Bristol, UK
Frederic Cazals	Sophia Antipolis, France
CQ Chang	University of Hong Kong, China
Marco Chierici	Bruno Kessler Foundation, Trento, Italy
Theo Damoulas	University of Glasgow, UK
Richard Edwards	University of Southampton, UK
Maurizio Filippone	University of Sheffield, UK
Alexandru Floares	Oncological Institute Cluj, Romania
Jennifer Hallinan	University of Newcastle, UK
Jin-Kao Hao	University of Angers, France
Jaap Heringa	VU University Amsterdam, The Netherlands
Antti Honkela	Helsinki University of Technology, Finland
Giuseppe Jurman	Bruno Kessler Foundation, Trento, Italy
R. Krishna Murthy Karuturi	Genome Institute of Singapore
Samuel Kaski	Helsinki University of Technology, Finland
Alex Kochetov	Russian Academy of Sciences, Russia
Mehmet Koyuturk	Case Western Reserve University, Cleveland, USA
Zoe Lacroix	Arizona State University, USA
Tak-Wah Lam	University of Hong Kong, China
Kee Khoon Lee	Institute of High Performance Computing, Singapore
Pietro Liò	University of Cambridge, UK
Xuejun Liu	Nanjing University of Aeronautics and Astronautics, China
Francesco Masulli	University of Genova, Italy
Mariofanna Milanova	Donaghey College of Engineering and Information Technology, USA
Sach Mukherjee	University of Warwick, UK

Alioune Ngom	University of Windsor, Canada
Carlotta Orsenigo	Politecnico di Milano, Italy
Nikhil Pal	Indian Statistical Institute, India
Magnus Rattray	University of Manchester, UK
Simon Rogers	University of Glasgow, UK
Juho Rousu	University of Helsinki, Finland
Anastasia Samsonova	Harvard University, USA
Alexander Schliep	Max Planck Institute for Molecular Genetics, Berlin, Germany
Roberto Tagliaferri	University of Salerno, Italy
Gwenn Volkert	Kent State University, USA
David Wild	University of Warwick, UK
Hong Yan	City University of Hong Kong, China
Jing Yang	Qingdao Institute of Bioenergy and Bioprocess Technology, China
Yan-Qing Zhang	Georgia State University, USA

Conference Organizing Committee

Conference Chairs

Visakan Kadirkamanathan	University of Sheffield, UK
Guido Sanguinetti	University of Sheffield, UK

General Co-chairs

Raj Acharya	PennState, USA
Madhu Chetty	Monash University, Australia
Jagath Rajapakse	Nanyang Technological University, Singapore

Program Chairs

Mahesan Niranjan	University of Southampton, UK
Mark Girolami	University of Glasgow, UK

Tutorial Chair

Florence d'Alché-Buc	University of Évry, France
----------------------	----------------------------

Special Sessions Chair

Cesare Furlanello	Fondazione Bruno Kessler, Italy
-------------------	---------------------------------

Publicity Chair

Elena Marchiori	Radboud University of Nijmegen, The Netherlands
-----------------	--

Publications Chair

Josselin Noirel University of Sheffield, UK

Local Organization Chair

Daniel Coca University of Sheffield, UK

Finance Chair

Andrew Zammit Mangion University of Sheffield, UK

Webmaster

Maurizio Filippone University of Sheffield, UK

Table of Contents

Evolutionary Parameters in Sequence Families: Cold Adaptation of Enzymes	1
<i>Said Hassan Ahmed and Tor Flå</i>	
MProfiler: A Profile-Based Method for DNA Motif Discovery	13
<i>Doaa Altarawy, Mohamed A. Ismail, and Sahar M. Ghanem</i>	
On Utilizing Optimal and Information Theoretic Syntactic Modeling for Peptide Classification	24
<i>Eser Ayy�n, B. John Oommen, and Zehra Cataltepe</i>	
Joint Tracking of Cell Morphology and Motion	36
<i>Jierong Cheng, Esther G.L. Koh, Sohail Ahmed, and Jagath C. Rajapakse</i>	
Multiclass Microarray Gene Expression Analysis Based on Mutual Dependency Models	46
<i>Girija Chetty and Madhu Chetty</i>	
An Efficient Convex Nonnegative Network Component Analysis for Gene Regulatory Network Reconstruction	56
<i>Jisheng Dai, Chunqi Chang, Zhongfu Ye, and Yeung Sam Hung</i>	
Using Higher-Order Dynamic Bayesian Networks to Model Periodic Data from the Circadian Clock of <i>Arabidopsis Thaliana</i>	67
<i>R�n�n Daly, Kieron D. Edwards, John S. O'Neill, Stuart Aitken, Andrew J. Millar, and Mark Girolami</i>	
Sequential Hierarchical Pattern Clustering	79
<i>Bassam Farran, Amirthalingam Ramanan, and Mahesan Niranjana</i>	
Syntactic Pattern Recognition Using Finite Inductive Strings	89
<i>Paul Fisher, Howard Fisher, Jinsuk Baek, and Cleopas Angaye</i>	
Evidence-Based Clustering of Reads and Taxonomic Analysis of Metagenomic Data	102
<i>Gianluigi Folino, Fabio Gori, Mike S.M. Jetten, and Elena Marchiori</i>	
Avoiding Spurious Feedback Loops in the Reconstruction of Gene Regulatory Networks with Dynamic Bayesian Networks	113
<i>Marco Grzegorzcz�k and Dirk Husmeier</i>	

Ligand Electron Density Shape Recognition Using 3D Zernike Descriptors	125
<i>Prasad Gunasekaran, Scott Grandison, Kevin Cowtan, Lora Mak, David M. Lawson, and Richard J. Morris</i>	
Definition of Valid Proteomic Biomarkers: A Bayesian Solution	137
<i>Keith Harris, Mark Girolami, and Harald Mischak</i>	
Inferring Meta-covariates in Classification	150
<i>Keith Harris, Lisa McMillan, and Mark Girolami</i>	
A Multiobjective Evolutionary Algorithm for Numerical Parameter Space Characterization of Reaction Diffusion Systems	162
<i>Tim Hohm and Eckart Zitzler</i>	
Knowledge-Guided Docking of WW Domain Proteins and Flexible Ligands	175
<i>Haiyun Lu, Hao Li, Shamima Banu Bte Sm Rashid, Wee Kheng Leow, and Yih-Cherng Liou</i>	
Distinguishing Regional from Within-Codon Rate Heterogeneity in DNA Sequence Alignments	187
<i>Alexander V. Mantzaris and Dirk Husmeier</i>	
A Hybrid Metaheuristic for Biclustering Based on Scatter Search and Genetic Algorithms	199
<i>Juan A. Nepomuceno, Alicia Troncoso, and Jesús S. Aguilar-Ruiz</i>	
Di-codon Usage for Gene Classification	211
<i>Minh N. Nguyen, Jianmin Ma, Gary B. Fogel, and Jagath C. Rajapakse</i>	
Counting Patterns in Degenerated Sequences	222
<i>Grégory Nuel</i>	
Modelling Stem Cells Lineages with Markov Trees	233
<i>Victor Olariu, Daniel Coca, Stephen A. Billings, and Visakan Kadiramanathan</i>	
Bi-clustering of Gene Expression Data Using Conditional Entropy	244
<i>Afolabi Olomola and Sumeet Dua</i>	
c-GAMMA: Comparative Genome Analysis of Molecular Markers	255
<i>Pierre Peterlongo, Jacques Nicolas, Dominique Lavenier, Raoul Vorc'h, and Joël Querellou</i>	
Semi-supervised Prediction of Protein Interaction Sentences Exploiting Semantically Encoded Metrics	270
<i>Tamara Polajnar and Mark Girolami</i>	

Classification of Protein Interaction Sentences via Gaussian Processes	282
<i>Tamara Polajnar, Simon Rogers, and Mark Girolami</i>	
MCMC Based Bayesian Inference for Modeling Gene Networks	293
<i>Ramesh Ram and Madhu Chetty</i>	
Efficient Optimal Multi-level Thresholding for Biofilm Image Segmentation	307
<i>Darío Rojas, Luis Rueda, Homero Urrutia, and Alioune Ngom</i>	
A Pattern Classification Approach to DNA Microarray Image Segmentation	319
<i>Luis Rueda and Juan Carlos Rojas</i>	
Drugs and Drug-Like Compounds: Discriminating Approved Pharmaceuticals from Screening-Library Compounds	331
<i>Amanda C. Schierz and Ross D. King</i>	
Fast SCOP Classification of Structural Class and Fold Using Secondary Structure Mining in Distance Matrix	344
<i>Jian-Yu Shi and Yan-Ning Zhang</i>	
Short Segment Frequency Equalization: A Simple and Effective Alternative Treatment of Background Models in Motif Discovery	354
<i>Kazuhito Shida</i>	
Bayesian Optimization Algorithm for the Non-unique Oligonucleotide Probe Selection Problem	365
<i>Laleh Soltan Ghoraie, Robin Gras, Lili Wang, and Alioune Ngom</i>	
Microarray Time-Series Data Clustering via Multiple Alignment of Gene Expression Profiles	377
<i>Numanul Subhani, Alioune Ngom, Luis Rueda, and Conrad Burden</i>	
Recursive Neural Networks for Undirected Graphs for Learning Molecular Endpoints	391
<i>Ian Walsh, Alessandro Vullo, and Gianluca Pollastri</i>	
Enhancing the Effectiveness of Fingerprint-Based Virtual Screening: Use of Turbo Similarity Searching and of Fragment Frequencies of Occurrence	404
<i>Shereena M. Arif, Jérôme Hert, John D. Holliday, Nurul Malim, and Peter Willett</i>	
Patterns, Movement and Clinical Diagnosis of Abdominal Adhesions ...	415
<i>Benjamin Wright, John Fenner, Richard Gillott, Paul Spencer, Patricia Lawford, and Karna Dev Bardhan</i>	

Class Prediction from Disparate Biological Data Sources Using an Iterative Multi-Kernel Algorithm.....	427
<i>Yiming Ying, Colin Campbell, Theodoros Damoulas, and Mark Girolami</i>	
Cross-Platform Analysis with Binarized Gene Expression Data.....	439
<i>Salih Tuna and Mahesan Niranjan</i>	
Author Index	451