Pattern Recognition in Bioinformatics

6th IAPR International Conference, PRIB 2011
Delft, The Netherlands, November 2-4, 2011
Proceedings

Springer
In modern biology, high-throughput measurement devices allow life scientists to gather data at unprecedented rates. To make sense of these data, computational biologists and system biologists construct quantitative models, many of which depend on pattern recognition techniques. Application of these techniques is challenging due to the large volumes of data and background information, noisy measurements and target outputs, highly diverse data types etc. To address these bioinformatics challenges, research in pattern recognition has been forced to move beyond the simple application of classifiers to high-throughput data, with which the field started over a decade ago. In 2006, the Pattern Recognition in Bioinformatics (PRIB) meeting series was established, under the auspices of the International Association of Pattern Recognition (IAPR), to support the exchange of ideas on how pattern recognition can be developed and extended to support the life sciences.

The Sixth conference in this series was organized in Delft, The Netherlands, November 2–4, 2011. In total, 35 papers were submitted, of which 29 were selected for publication in these proceedings after peer review by the Program Committee. The contributions demonstrate the wide range of possible applications of pattern recognition in bioinformatics: novel algorithms to handle traditional pattern recognition problems such as (bi)clustering, classification and feature selection; applications of (novel) pattern recognition techniques to infer and analyze biological networks and studies on specific problems such as biological image analysis and the relation between sequence and structure. In addition to the 29 paper presentations, the conference saw invited presentations by Yves Moreau (Bioinformatics Research Group, Katholieke Universiteit Leuven, Belgium), Mark Girolami (Department of Statistical Science, University College London, UK), Pierre Baldi (Institute of Genomics and Bioinformatics, University of California, Irvine, USA), Alvis Brazma (European Bioinformatics Institute, Hinxton, UK) and Berend Snel (Department of Biology, Utrecht University, The Netherlands).

We would like to thank all participants in the conference, in particular the authors who contributed their work to this book. We are grateful for the support of IAPR’s Technical Committee 20 and the PRIB Steering Committee. The help of the members of the Program Committee was invaluable in evaluating the submissions and defining the final program. EasyChair (http://www.easychair.org) was of great help in handling the submissions, reviews, decisions, and accepted papers. We would also like to thank the co-organizers (Jeroen de Ridder, Manuele Bicego, David Tax and Thomas Abeel) as well as members of the Pattern Recognition & Bioinformatics group (Delft University of Technology) for their
assistance before and during the conference as well as the Netherlands Bioinformatics Centre and the EU PASCAL2 Network of Excellence for financial support. Last but not least, the conference could not have been organized without the invaluable help of Saskia Peters, who took care of administration and local arrangements.

August 2011

Marco Loog
Lodewyk Wessels
Marcel Reinders
Dick de Ridder
Organization

Program Committee

Jesús S. Aguilar-Ruiz, Pablo de Olavide University, Spain
Shandar Ahmad, National Institute of Biomedical Innovation, Japan
Tatsuya Akutsu, Kyoto University, Japan
Jaume Bacardit, University of Nottingham, UK
Rainer Breitling, University of Groningen, The Netherlands
Sebastian Böcker, Friedrich Schiller University Jena, Germany
Frederic Cazals, INRIA Sophia, France
Dick De Ridder, Delft University of Technology, The Netherlands
Tjeerd Dijkstra, Radboud University Nijmegen, The Netherlands
Federico Divina, Pablo de Olavide University, Spain
Bas E. Dutilh, CMBI / NCMLS / UMCN, The Netherlands
Richard Edwards, University of Southampton, UK
Maurizio Filippone, University College London, UK
Rosalba Giugno, University of Catania, Italy
Michael Habeck, MPIs for Biological Cybernetics and Developmental Biology, Germany
Jin-Kao Hao, University of Angers, France
Morihiro Hayashida, Kyoto University, Japan
Tom Heskes, Radboud University Nijmegen, The Netherlands
Pavol Jancura, Radboud University Nijmegen, The Netherlands
Zhenyu Jia, University of California, Irvine, USA
Giuseppe Jurman, Fondazione Bruno Kessler, Italy
Visakan Kadirkamanathan, The University of Sheffield, UK
Seyoung Kim, Carnegie Mellon University, USA
Walter Kosters, Leiden University, The Netherlands
Xuejun Liu, Nanjing University of Aeronautics and Astronautics, China
Stefano Lonardi, UC Riverside, USA
Elena Marchiori, Radboud University, The Netherlands
Francesco Masulli, University of Genoa, Italy
Jason Moore, Dartmouth College, USA
Alison Motsinger-Reif, North Carolina State University, USA
VIII Organization

Vadim Mottl Computing Center of the Russian Academy of Sciences, Russia
Sach Mukherjee University of Warwick, UK
Tamás Nepusz University of London, UK
Emrah Nikerel Delft University of Technology, The Netherlands
Josselin Noirel University of Sheffield, UK
Carlotta Orsenigo Politecnico di Milano, Italy
Andrea Passerini University of Trento, Italy
Thang Pham VU University Medical Center, The Netherlands
Esa Pitkänen University of Helsinki, Finland
Clara Pizzuti Institute for High Performance Computing and Networking, Italy
Beatriz Pontes University of Seville, Spain
Marylyn Ritchie Vanderbilt University, USA
Miguel Rocha University of Minho, Portugal
Simon Rogers Department of Computing Science
Juho Rousu University of Helsinki, Finland
Yvan Saeys Flanders Institute for Biotechnology, Belgium
Guido Sanguinetti University of Edinburgh, UK
Jun Sese Ochanomizu University, Japan
Evangelos Simeonidis Manchester Centre for Integrative Systems Biology, UK
Johan Suykens K.U. Leuven, Belgium
Roberto Tagliaferri University of Salerno, Italy
Alexey Tsymbal Siemens AG, Germany
Bastiaan Van Den Berg Delft University of Technology, The Netherlands
Lodewyk Wessels NKI-AVL, The Netherlands
Wynand Winterbach Delft University of Technology, The Netherlands
Hong Yan City University of Hong Kong, SAR China
Haixuan Yang Royal Holloway University of London, UK
# Table of Contents

## Session 1: Clustering

A New Framework for Co-clustering of Gene Expression Data .......................... 1  
*Shuzhong Zhang, Kun Wang, Bilian Chen, and Xiuzhen Huang*

Biclustering of Expression Microarray Data Using Affinity Propagation .................................................................................................................. 13  
*Alessandro Farinelli, Matteo Denitto, and Manuele Bicego*

A Two-Way Bayesian Mixture Model for Clustering in Metagenomics ... 25  
*Shruthi Prabhakara and Raj Acharya*

CRiSPy-CUDA: Computing Species Richness in 16S rRNA Pyrosequencing Datasets with CUDA ......................................................... 37  
*Zejun Zheng, Thuy-Diem Nguyen, and Bertil Schmidt*

## Session 2: Biomarker Selection and Classification (1)

New Gene Subset Selection Approaches Based on Linear Separating Genes and Gene-Pairs ................................................................. 50  
*Amirali Jafarian, Alioune Ngom, and Luis Rueda*

Identification of Biomarkers for Prostate Cancer Prognosis Using a Novel Two-Step Cluster Analysis .................................................. 63  
*Xin Chen, Shizhong Xu, Yipeng Wang, Michael McClelland,  
Zhengyu Jia, and Dan Mercola*

Renal Cancer Cell Classification Using Generative Embeddings and Information Theoretic Kernels .................................................. 75  
*Manuele Bicego, Aydin Ula¸s, Peter Schüßler, Umberto Castellani,  
Vittorio Murino, André Martins, Pedro Aguiar, and  
Mario Figueiredo*

## Session 3: Network Inference and Analysis (1)

Integration of Epigenetic Data in Bayesian Network Modeling of Gene Regulatory Network ................................................................. 87  
*Jie Zheng, Iti Chaturvedi, and Jagath C. Rajapakse*

Metabolic Pathway Inference from Time Series Data: A Non Iterative Approach ................................................................. 97  
*Laura Astola, Marian Groenenboom, Victoria Gomez Roldan,  
Fred van Eeuwijk, Robert D. Hall, Arnaud Bovy, and Jaap Molenaar*
Highlighting Metabolic Strategies Using Network Analysis over Strain Optimization Results .......................................................... 109  
José Pedro Pinto, Isabel Rocha, and Miguel Rocha

Session 4: Biomarker Selection and Classification (2)

Wrapper- and Ensemble-Based Feature Subset Selection Methods for Biomarker Discovery in Targeted Metabolomics ...................... 121  
Holger Franken, Rainer Lehmann, Hans-Ulrich Häring, Andreas Fritsche, Norbert Stefan, and Andreas Zell

Ensemble Logistic Regression for Feature Selection ...................... 133  
Roman Zakharov and Pierre Dupont

Gene Selection in Time-Series Gene Expression Data .................... 145  
Prem Raj Adhikari, Bimal Babu Upadhyyaya, Chen Meng, and Jaakko Hollmén

Multi-task Drug Bioactivity Classification with Graph Labeling Ensembles ................................................................. 157  
Hongyu Su and Juho Rousu

Session 5: Image Analysis (1)

A Bilinear Interpolation Based Approach for Optimizing Hematoxylin and Eosin Stained Microscopical Images ......................... 168  
Kaya Kuru and Sertan Girgin

Automatic Localization of Interest Points in Zebrafish Images with Tree-Based Methods ......................................................... 179  
Olivier Stern, Raphaël Marée, Jessica Aceto, Nathalie Jeanray, Marc Muller, Louis Wehenkel, and Pierre Geurts

Session 6: Biomarker Selection and Classification (3)

A New Gene Selection Method Based on Random Subspace Ensemble for Microarray Cancer Classification ................................. 191  
Giuliano Armano, Camelia Chira, and Nima Hatami

A Comparison on Score Spaces for Expression Microarray Data Classification ............................................................ 202  
Alessandro Perina, Pietro Lovato, Marco Cristani, and Manuele Bicego
Flux Measurement Selection in Metabolic Networks ........................... 214
Wout Megchelenbrink, Martijn Huynen, and Elena Marchiori

Session 7: Network Inference and Analysis (2)

Lagrangian Relaxation Applied to Sparse Global Network Alignment ... 225
Mohammed El-Kebir, Jaap Heringa, and Gunnar W. Klau

Stability of Inferring Gene Regulatory Structure with Dynamic
Bayesian Networks ............................................................. 237
Jagath C. Rajapakse and Iti Chaturvedi

Integrating Protein Family Sequence Similarities with Gene Expression
to Find Signature Gene Networks in Breast Cancer Metastasis .......... 247
Sepideh Babaei, Erik van den Akker, Jeroen de Ridder, and
Marcel Reinders

Session 8: Sequence, Structure, and Interactions

Estimating the Class Posterior Probabilities in Protein Secondary
Structure Prediction .............................................................. 260
Yann Guermeur and Fabienne Thomarat

Shape Matching by Localized Calculations of Quasi-Isometric Subsets,
with Applications to the Comparison of Protein Binding Patches ...... 272
Frédéric Cazals and Noël Malod-Dognin

Using Kendall-τ Meta-bagging to Improve Protein-Protein Docking
Predictions ................................................................. 284
Jérôme Azé, Thomas Bourquard, Sylvie Hamel, Anne Poupon, and
David W. Ritchie

PAAA: A Progressive Iterative Alignment Algorithm Based
on Anchors ................................................................. 296
Ahmed Mokaddem and Mourad Elloumi

Session 9: Image Analysis (2)

Heat Diffusion Based Dissimilarity Analysis for Schizophrenia
Classification ................................................................. 306
Aydın Ulaş, Umberto Castellani, Vittorio Murino, Marcella Bellani,
Michele Tansella, and Paolo Brambilla
Epithelial Area Detection in Cytokeratin Microscopic Images Using MSER Segmentation in an Anisotropic Pyramid ....................... 318
Cristian Smochina, Radu Rogojanu, Vasile Manta, and Walter Kropatsch

Pattern Recognition in High-Content Cytomics Screens for Target Discovery – Case Studies in Endocytosis .......................... 330
Lu Cao, Kuan Yan, Leah Winkel, Marjo de Graauw, and Fons J. Verbeek

Author Index .......................................................... 343