

Lecture Notes in Artificial Intelligence 3303

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Jesús A. López Emilio Benfenati
Werner Dubitzky (Eds.)

Knowledge Exploration in Life Science Informatics

International Symposium KELSI 2004
Milan, Italy, November 25-26, 2004
Proceedings

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Jesús A. López
University of Southern Queensland
Department of Mathematics and Computing, Faculty of Sciences
Toowoomba, Queensland, 4350 Australia
E-mail: lopez@usq.edu.au

Emilio Benfenati
Istituto di Ricerche Farmacologiche "Mario Negri"
(Laboratory of Environmental Chemistry and Toxicology)
Via Eritrea 62, 20157 Milano, Italy
E-mail: benfenati@marionegri.it

Werner Dubitzky
University of Ulster
School of Biomedical Sciences
Cromore Road, Coleraine BT 52 ISA, Northern Ireland
E-mail: w.dubitzky@ulster.ac.uk

Library of Congress Control Number: 2004115081

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

ISSN 0302-9743

ISBN 3-540-23927-8 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11358725 06/3142 5 4 3 2 1 0

Preface

This volume of the Springer Lecture Notes in Computer Science series contains the contributions presented at the International Symposium on Knowledge Exploration in Life Science Informatics (KELSI 2004) held in Milan, Italy, 25–26 November 2004. The two main objectives of the symposium were:

- To explore the symbiosis between information and knowledge technologies and various life science disciplines, such as biochemistry, biology, neuroscience, medical research, social sciences, and so on.
- To investigate the synergy among different life science informatics areas, including cheminformatics, bioinformatics, neuroinformatics, medical informatics, systems biology, socionics, and others.

Modern life sciences investigate phenomena and systems at the level of molecules, cells, tissues, organisms, and populations. Typical areas of interest include natural evolution, development, disease, behavior, cognition, and consciousness. This quest is generating an overwhelming and fast-growing amount of data, information, and knowledge, reflecting living systems at different levels of organization. Future progress of the life sciences will depend on effective and efficient management, sharing, and exploitation of these resources by computational means.

Life science informatics is fast becoming a generic and overarching information technology (IT) discipline for the life sciences. It includes areas such as cheminformatics, bioinformatics, neuroinformatics, medical informatics, socionics, and others. While the precise scientific questions and goals differ within the various life science disciplines, there is a considerable overlap in terms of the required key IT methodologies and infrastructures. Critical technologies include *databases*, *information bases* (i.e., containing aggregated, consolidated, derived data), *executable models* (i.e., knowledge-based and simulation systems), and *emerging grid computing* infrastructures and systems (facilitating seamless sharing and interoperation of widely dispersed computational resources and organizations). These base technologies are complemented by a range of enabling methodologies and systems such as knowledge management and discovery, data and text mining, machine learning, intelligent systems, artificial and computational intelligence, human-computer interaction, computational creativity, knowledge engineering, artificial life, systems science, and others.

This symposium was a first step towards investigating the synergy of these knowledge and information technologies across a wide range of life science disciplines.

Milan, Italy, November 2004

Jesús A. López
Emilio Benfenati
Werner Dubitzky

Acknowledgments

KELSI 2004 was organized and sponsored by the European Science Foundation's COST Action 282. We thank all Action 282 members for their support, inspiration, and contributions. Special thanks goes to the invited speakers – Prof. Michal Linial, Prof. Gustavo Deco, Prof. Stephen Muggleton, and Dr. Martin Kuiper – who reminded us of the mind-boggling breadth and depth of modern life science informatics and the challenges involved. Finally we would like to extend our gratitude to the members of the KELSI 2004 International Program Committee and the reviewers: Kevin Cohen (University of Colorado Health Sciences Center), Joerg Hakenberg (Humboldt-Universität zu Berlin), Sanjay Shamara Kadam and Valadi Krishnamoorthy Jayaraman (both Centre for Development of Advanced Computing, Pune, India), David McSherry (University of Ulster), Alessandra Roncaglioni (Istituto di Ricerche Farmacologiche – “Mario Negri”), and Alexander Seewald (Austrian Research Institute for Artificial Intelligence).

International Program Committee

Jesús A. López (IPC Co-chair), University of Southern Queensland, Australia
Emilio Benfenati (IPC Co-chair and Local Organizer),

Istituto di Ricerche Farmacologiche – “Mario Negri”, Milan, Italy

Werner Dubitzky (IPC Co-chair), University of Ulster, Coleraine, UK

Agnar Aamodt, Norwegian University of Science and Technology (NTNU),
Trondheim, Norway

Michael R. Berthold, University of Konstanz, Germany

Leon Bobrowski, Institute of Biocybernetics and Biomedical Engineering,
Warsaw, Poland

Eric G. Bremer, Children's Memorial Hospital, Chicago, IL, USA

Rui Meireles de Brito, University of Coimbra, Coimbra, Portugal

Amílcar Cardoso, CISUC, University of Coimbra, Coimbra, Portugal

Gustavo Deco, Institució Catalana de Recerca i Estudis Avançats (ICREA),
Barcelona, Spain

Catherine DeSesa, SPSS Inc., Chicago, IL, USA

Moustafa M. Ghanem, Imperial College, London, UK

Geerd H. Diercksen, Max Planck Institute for Astrophysics, Garching, Germany

Robin Gras, Proteome Informatics Group, Swiss Institute of Bioinformatics,
Switzerland

Des Higgins, University College Cork, Ireland

Chun-Hsi Huang, University of Connecticut, Storrs, USA

Seiya Imoto, University of Tokyo, Japan

Igor Jurisica, Ontario Cancer Institute, Canada

Astrid Lægreid, Norwegian University of Science and Technology, Trondheim, Norway

Michael Liebman, University of Pennsylvania, Philadelphia, USA

Stephanie McKeown, University of Ulster, UK

Alain Mille, Université Claude Bernard Lyon 1, LIRIS, France

Ray C. Paton, University of Liverpool, UK

Enric Plaza, IIIA-CSIC, Catalonia, Spain

Mathilde Romberg, Research Center Jülich, Jülich, Germany

Vlado Stankovski, University of Ljubljana, Slovenia

Vijayraghavan Sundararajan, Centre for Development of Advanced Computing,
Pune, India

Brigitte Trousse, INRIA, AxIS Project Team, Sophia-Antipolis, France

Danny van Welden, University of Gent, Belgium

Gerhard Widmer, Austrian Research Institute for Artificial Intelligence, Vienna, Austria

Table of Contents

A Pen-and-Paper Notation for Teaching Biosciences	1
<i>Johannes J. Mandel and Niall M. Palfreyman</i>	
An Exploration of Some Factors Affecting the Correlation of mRNA and Proteomic Data	9
<i>Catherine J. Hack and Jesús A. López</i>	
Improving Rule Induction Precision for Automated Annotation by Balancing Skewed Data Sets	20
<i>Gustavo E.A.P.A. Batista, Maria C. Monard, and Ana L.C. Bazzan</i>	
A Randomized Algorithm for Distance Matrix Calculations in Multiple Sequence Alignment	33
<i>Sanguthevar Rajasekaran, Vishal Thapar, Hardik Dave, and Chun-Hsi Huang</i>	
Extracting Sequential Patterns for Gene Regulatory Expressions Profiles	46
<i>Doru Tanasa, Jesús A. López, and Brigitte Trousse</i>	
Data Analysis of Microarrays Using SciCraft	58
<i>Bjørn K. Alsberg, Lars Kirkhus, Truls Tangstad, and Endre Anderssen</i>	
Functional Data Analysis of the Dynamics of Gene Regulatory Networks	69
<i>Tomohiro Ando, Seiya Imoto, and Satoru Miyano</i>	
Text Mining of Full Text Articles and Creation of a Knowledge Base for Analysis of Microarray Data	84
<i>Eric G. Bremer, Jeyakumar Natarajan, Yonghong Zhang, Catherine DeSesa, Catherine J. Hack, and Werner Dubitzky</i>	
Analysis of Protein/Protein Interactions Through Biomedical Literature: Text Mining of Abstracts vs. Text Mining of Full Text Articles	96
<i>Eric P.G. Martin, Eric G. Bremer, Marie-Claude Guerin, Catherine DeSesa, and Olivier Jouve</i>	
Ranking for Medical Annotation: Investigating Performance, Local Search and Homonymy Recognition	109
<i>Alexander K. Seewald</i>	
A New Artificial Life Formalization Model: A Worm with a Bayesian Brain	124
<i>Fidel Aznar Gregori, Maria Del Mar Pujol López, Ramón Rizo Aldeguez, and Pablo Suau Pérez</i>	
Teaching Grasping to a Humanoid Hand as a Generalization of Human Grasping Data	139
<i>Michele Folgheraiter, Ilario Baragiola, and Giuseppina Gini</i>	

JavaSpaces – An Affordable Technology for the Simple Implementation of Reusable Parallel Evolutionary Algorithms	151
<i>Christian Setzkorn and Ray C. Paton</i>	
Detecting and Adapting to Concept Drift in Bioinformatics	161
<i>Michaela Black and Ray Hickey</i>	
Feature Extraction and Classification of the Auditory Brainstem Response Using Wavelet Analysis	169
<i>Rui Zhang, Gerry McAllister, Bryan Scotney, Sally McClean, and Glen Houston</i>	
Evaluation of Outcome Prediction for a Clinical Diabetes Database	181
<i>Yue Huang, Paul McCullagh, Norman Black, and Roy Harper</i>	
Cytochrome P450 Classification of Drugs with Support Vector Machines Implementing the Nearest Point Algorithm	191
<i>Achim Kless and Tatjana Eitrich</i>	
Multiple-Instance Case-Based Learning for Predictive Toxicology	206
<i>Eva Armengol and Enric Plaza</i>	
Modelling and Prediction of Toxicity of Environmental Pollutants	221
<i>Frank Lemke, Johann-Adolf Müller, and Emilio Benfenati</i>	
Modelling Aquatic Toxicity with Advanced Computational Techniques: Procedures to Standardize Data and Compare Models	235
<i>Emilio Benfenati</i>	
Author Index	249